NUCLEAR MAGNETIC RESONANCE CHARACTERIZATION OF AN UNUSUAL 37 KILODALTON NONRIBOSOMAL PEPTIDE SYNTHETASE EPIMERIZATION DOMAIN

by
Scott R. Nichols

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
June 2016

Abstract

Yersiniabactin synthetase is a megadalton nonribosomal peptide synthetase that produces the iron-chelator yersiniabactin, which is essential for parasitic growth of *Yersinia pestis*, the etiological agent of plague. Yersiniabactin synthetase employs domains that reflect a departure from classical non-ribosomal peptide synthetase assembly-line machinery including an unusual epimerization domain. Embedded in the primary sequence of an adenylation domain, which selects one substrate to be joined to the growing nonribosomal peptide chain, the epimerization domain controls the stereochemistry of an essential chiral center of yersiniabactin. This thesis describes characterization of this epimerization domain using nuclear magnetic resonance spectroscopy. We were able to define clear domain boundaries, express and purify the epimerization domain, and assign ~80% of backbone resonances. We used non-uniform sampling and partial deuteration to dramatically increase sensitivity in detection of the 37 kDa protein. NMR data spectral complexity was overcome with novel strategies in covariance nuclear magnetic resonance. An equilibrium protein refolding strategy (in protonated solvent) hastened slow back exchange of amide ND for NH. Finally, a novel TS-(HN-TROSY/HC-HMQC)-NOESY experiment is described for collection of NOESY distance constraints on a uniformly ^{13}C ^{15}N ^{2}H labeled protein that is selectively protonated at the methyl groups of isoleucine, leucine, and valine.

Advisor:	Professor Dominique P. Frueh, Ph.D.
Thesis Reader:	Professor Joel R. Tolman, Ph.D.

Acknowledgements

This thesis is dedicated in loving memory to Professor Theresa L. Gioannini, Ph.D., my first mentor whose brilliance was surpassed only by her magnanimity. Both Theresa and her husband, Professor Jerrold P. Weiss, Ph.D. provided me with a strong scientific foundation and empowered me to chase my own ideas fiercely. Their love and support have carried me on my scientific journey. I would also like to thank members of the University of Iowa Inflammation Program, specifically Dr. Polonca Prohinar, Desheng Zheng, Dr. Ramona L. McCaffrey, and Dr. Anne N. Shemon for teaching me experimental technique and design. I thank Professor Adrian H. Elcock, Ph.D. of the University of Iowa for introducing me to computational biophysics. Adrian challenged me to think quantitatively and to choose hard, audacious projects and to hold others and myself to a rigorous standard.

I extend my deep appreciation to the members of the Unité Toxines et Pathogénie Bactérienne at the Pasteur Institute in Paris, France. My mentors Dr. Lhousseine Touqui and Dr. Pierre L. Goossens taught me Microbiology and Immunology techniques that spanned *in vitro* biochemistry to *in vivo* animal models. I am indebted to Dr. Fabien P. Dumetz for his scientific insight and friendship.

I thank the laboratories of Professor Wolfgang Peti, Ph.D. and Professor Rebecca Page, Ph.D. of Brown University where I learned the foundations of structural biology. My time in the Page-Peti Lab prepared me extremely well for graduate school, and Wolfgang and Rebecca have remained constantly supportive throughout my career. I thank my good friend Professor Anderson de Sá Pinheiro,

Ph.D. for his love and support over the past decade, as well as Dr. David A. Critton and Dr. Dana M. Francis for their personal and scientific advice throughout the years.

I am extremely thankful to Dr. Ananya Majumdar who has been both a valuable mentor and a trusted friend over the years. Always patient and kind, Ananya has eased the most difficult parts of graduate school, both technically and emotionally. I am grateful to the other members of my thesis committee for support and expertise: Professors Joel Tolman, Ph.D., Craig A. Townsend, Ph.D., and James T. Stivers, Ph.D. I thank my lab mates, Dr. Andrew C. Goodrich and Dr. Subrata H. Mishra for their insights and feedback and contribution to a stimulating environment both personally and scientifically. I am profoundly thankful to my best friend and soon-to-be Dr. Bradley J. Harden who taught me about mathematics, programming, and aspects of quantum mechanics that were indispensible for my research. Bradley has been unwavering in his support and a true friend whose generosity knows no bounds. Most importantly, I thank my adviser Professor Dominique P. Frueh, Ph.D. for his mentorship over the years. I deeply respect his focus, determination, and meticulousness. I have learned to properly define problems and to choose the right tools to solve them. Dominique refuses incomplete answers and seeks to understand how every single datum fits into the overall picture. I am confident that this approach will serve me in all my future endeavors, and for this, I am endlessly grateful.

Finally, I am fortunate to be surrounded by a family and network of friends who love and support me. They soften moments of pain, and they make the victories brighter. I am blessed to have become so close to Nathan Klein in the past year and look forward to our future together. My parents, Jane and Leo, and my brother, Eric, show me every day that love survives geography.

"For science is more than the search for truth, more than a challenging game, more than a profession. It is a life that a diversity of people lead together, in the closest proximity, a school for the social living. We are members, one of another."

-Professor Alexander George Ogston

Australian Biochemical Society Annual Lecutre, 1970

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction to High Molecular Weight Nuclear Magnetic Resonance Spectroscopy on Proteins

A modified portion of this text has been published in *eMagRes*, formerly *The Encyclopedia of NMR*.

Nuclear magnetic resonance (NMR) spectroscopy has a rich history in the fields of chemistry and physics, and it has become an indispensible tool for atomic-level investigations of biological macromolecules. Key to NMR's utility is its capacity to analyze complex biological macromolecules under physiological conditions. Although NMR spectroscopy is inherently insensitive, the NMR excited state is long-lived, and its spectroscopic lines are sharp. The long-lived NMR signal facilitates multi-dimensional spectroscopy, which in turn allows for generation of various correlation maps that allow for facile interrogation of biological macromolecules at the atomic-level. This introductory chapter briefly introduces the physical basis of NMR spectroscopy and multi-dimensional NMR spectroscopy, protein resonance assignment (of backbone and of side chain signals), and practical aspects for large and challenging proteins.

1. The physical basis of NMR spectroscopy

In NMR spectroscopy, the ground states and excited states do not pre-exist and need to be generated by application of a strong, external magnetic field. When nuclei of non-zero spin are placed into a strong external magnetic field, the energetic difference between the spin states increases in a manner proportional to the applied magnetic field (Zeeman splitting). The relationship between the magnetic field ($\hat{B}$ in Tesla, T) and energy of spin states (E in Joules, J) is described as

$$E = -\hat{\mu} \cdot \hat{B}$$

where $\hat{\mu}$ is the nuclear magnetic moment (in J T^{-1}). The dot product can be simplified by realizing that in an NMR spectrometer, the external magnetic field is aligned along z (in a vertical bore magnet).

$$\mu_z = -m\gamma\hbar$$

$$E = -m\gamma\hbar B_0 = -m\hbar\,\omega$$

Here, *m* is the magnetic quantum number, γ is the gyromagnetic ratio (a property of the nucleus, in rad S^{-1} T^{-1}), $\hbar$ is Plank's constant divided by 2π, and B_0 is the strength of the magnetic field (directed along *z*, in T). The selection rules for transitions in NMR spectroscopy require the change of spin states to be integer values:

$$\Delta m = m_{ES} - m_{GS} = \pm 1$$

Therefore, we can calculate the energy change for allowed transitions (between the excited state and ground state, ES and GS) in NMR spectroscopy:

$$\Delta E = E_{ES} - E_{GS} = \gamma\hbar B_0$$

When an ensemble of spins comes to equilibrium in an external magnetic field, the populations of states are described by the Boltzmann distribution. For spin ½ nuclei at 600 MHz (14.1 T) and 298 K, we observe that the two states, alpha and beta (ground state and excited state, respectively), are nearly equally populated. Approximately 1 in 10^5 spins are preferentially in the alpha state:

$$\frac{N_\beta}{N_\alpha} = \exp\left(\frac{-\Delta E}{k_B T}\right)$$

$$\frac{N_\beta}{N_\alpha} = \exp\left(\frac{-6.73\text{x}10^{7}\,\cancel{\text{rad T}^{-1}\text{s}^{-1}}\times 1.054\text{x}10^{-34}\,\cancel{\text{J s rad}^{-1}}\times 14.1\,\cancel{\text{T}}}{1.381\text{x}10^{-23}\,\cancel{\text{J K}^{-1}}\times 298.15\,\cancel{\text{K}}}\right) = 0.999976$$

In this above calculation, N_α and N_β refer to the relative numbers of spins in either the ground or excited states, respectively. This simple calculation reflects that NMR spectroscopy is extremely insensitive, especially when compared to other commonly used spectroscopies in molecular biophysics (namely, optical spectroscopies, such as ultraviolet and fluorescence). Transitions in NMR spectroscopy occur at very low energies, which are near thermal fluctuations. On the other hand, spontaneous emission, the process by which an excited state spontaneously returns to the ground state is inefficient in the case of NMR spectroscopy because the probability of spontaneous emission is proportional to the cube of the frequency of transition. Thus, the lifetime of the excited state is longer than those of other spectroscopies used to interrogate biological macromolecules. Secondly, the Heisenberg uncertainty principle with respect to energy and time describes that the relationship between the energy of transition and the lifetime of the excited state is inversely proportional.

$$\Delta E \times \Delta t \geq \frac{\hbar}{2}$$

Although NMR spectroscopy is among the least sensitive spectroscopies, the lifetimes of excited states are long-lived (on the order of milliseconds), which results in sharp spectral lines. Moreover, the long lifetime of the NMR excited state allows for the implementation of several complex multidimensional dimensional experiments. The investigations in this thesis and the methods applied to the system of study rely strongly upon the capacity to record multidimensional NMR experiments.

2. Multidimensional NMR spectroscopy

A. One-dimensional NMR spectroscopy

Beginning in the 1960's, NMR spectroscopy moved from continuous wave to pulse-acquire or Fourier Transform NMR spectroscopy. In continuous wave NMR spectroscopy, radiofrequency waves at specific frequencies were applied to samples such that essentially, only a single frequency is excited at once. By contrast, in Fourier Transform NMR spectroscopy, a single broad high power radiofrequency pulse of energy is applied to the sample. As a wide range of frequencies is affected, transitions occur between alpha and beta states of many spins.

To explain pulse NMR, I will employ the NMR experiment using the so-called vector model. Small differences in populations between energy levels of individual spin systems summed over an entire NMR sample, give rise to a bulk magnetization. The vector components of the bulk magnetization can be defined as:

$$\widehat{M} = \sum_{i=1}^{all\ spins} \mu_i$$

When a static magnetic field is applied to the sample, only the *z*-components of each magnetic dipole will produce magnetization (longitudinal magnetization, M_z) due to the slight difference in populations of spin. (M_0 is the magnitude of magnetization.)

$$M_z = M_0$$

In contrast, the *x*- and *y*-components of magnetization (transverse magnetization, M_x and M_y) are randomly distributed and sum to zero:

$$M_x = M_y = 0$$

In pulse NMR, a pulse of radiofrequency energy applies a torque on the bulk magnetization vector, which causes it to move from its longitudinal position (along +*z*, aligned with the magnetic field) into the transverse plane. Once the bulk magnetization vector moves into the transverse plane, it precesses at the Larmour Frequency, which in turn, induces a current (through Faraday's induction principle) in a highly susceptible coil that surrounds the sample. The rotating wave can be modeled as

$$s(t) = [\cos(\omega_0 t) + i\sin(\omega_0 t)]e^{-R_2 t}$$

where s(t) is the signal as a function of time, ω_0 is the Larmour frequency, and R_2 is the transverse relaxation rate (discussed in more detail below). The rotating wave signal is digitized and referred to as the Free Induction Decay (FID), which after Fourier Transform gives rise to a Lorentzian lineshape:

$$S(\omega) = \int_0^\infty s(t)e^{-i\omega t}\,dt = A(\omega) + iD(\omega)$$

$$A(\omega) = \frac{R_2}{(\omega_0 - \omega)^2 + (R_2)^2}$$

$$D(\omega) = \frac{(\omega_0 - \omega)}{(\omega_0 - \omega)^2 + (R_2)^2}$$

The center of the absorptive component (A(ω)) of the Lorentzian line of an individual nucleus is its chemical shift, which is given by:

$$\omega_0 = \gamma(1 - \sigma)B_0$$

in which σ is the chemical shielding factor, which describes how the local chemical (or electronic) environment around the nucleus affects its response to the applied magnetic field. The chemical shift of a nucleus is among the most information rich parameters in NMR spectroscopy. It is exquisitely sensitive to chemical environment, can be used to determine secondary structure, and can report on interactions of nuclei with other species (such as, ligands). The width of the Lorentzian line of an individual nucleus reflects the signal lifetime and is parameterized by the transverse relaxation rate constant, R_2. Slowly relaxing FIDs give rise to sharp Lorentzians. For a spin ½ nucleus, the relaxation rate that defines the width of a particular signal is dependent on field inhomogeneity and molecular motions that give rise to time-dependent fluctuations of dipole-dipole interactions and of the chemical shift anisotropy. These relaxation mechanisms will be discussed later in the context of the TROSY experiment. It is worthwhile to mention that R_2 becomes increasingly fast as molecular weight (and to a lesser degree magnetic field strength) increases; ergo, high molecular weight (slowly tumbling) molecules experience severe line broadening in NMR spectra.

When two NMR active nuclei are adjacent to one another in chemical structures, the lineshape of one nucleus can be "split" by the other nuclei. This "splitting" occurs because of so-called “hyperfine” interactions between NMR active nuclei and the electrons involved in covalent bonds between adjacent nuclei. In a two-spin system of spins I—S, in which both I and S are of spin ½, both spins I and S will appear as doublets in a 1D spectrum. The two lines centered at the chemical shift of I reflect the alpha and beta states of spin S, and the two lines centered at the chemical shift of spin S represent alpha and beta states of spin I. The distance between the two lines (interrogated at the frequencies of either I or S) equals J_{IS}, the coupling constant. (Also spin-spin coupling constant, J-coupling constant, or scalar coupling constant.)

B. Multidimensional NMR spectroscopy

In the one-dimensional NMR experiments, a single radiofrequency pulse puts a torque on the magnetization vector that pushes it into the transverse plane, where spins precess at their chemical shifts, giving rise to the FID. The collected signal varies as a function of acquisition time: S(t). In a two dimensional NMR experiment, the pulse sequence is constructed such that signals evolve as a function of a variable delay (t_1) within the pulse sequence as well as the acquisition time (t_2): $S(t_1,t_2)$. A matrix of data in (t_1, t_2) is collected: For a fixed value of (t_1) signals are collected for every point in t_2. Then, the value of t_1 is incremented and the experiment recollected for every point t_2. Upon completion of the experiment, the data matrix reflects signal as a function of both t_1 and t_2. When each FID of the entire matrix is subjected to Fourier Transform (along t_2), the signal intensity varies in each transient, and the data are of the form $S(t_1, \omega_2)$. Subsequent Fourier transform gives signals of the form

$S(\omega_1, \omega_2)$ which are themselves three dimensional (ω_1, ω_2, and intensity) and typically viewed as contour maps.

Two-dimensional correlation maps are nearly unique to NMR spectroscopy and provide access to information that few other techniques allow. Resonances that appear at the same frequency in both dimensions of homonuclear experiments are called "diagonal" peaks and those that appear at different frequencies in the two dimensions are "cross peaks." Depending on how the two dimensions are related to each other, cross-peaks can report on several different properties, but generally speaking, two-dimensional correlation maps reflect magnetization transfer. In these schemes, the chemical shift of one nucleus is encoded in an indirect dimension, then magnetization is transferred to another nucleus and the chemical shift of the second nucleus is encoded during the acquisition time. In NMR experiments magnetization transfer may occur through bond (through COSY-type transfers or through TOCSY-type transfers) or through space (through NOESY transfers). Each of these magnetization transfer schemes is discussed below since they are essential for understanding more complex multidimensional NMR experiments.

i) Magnetization transfers through COSY-type transfers

Magnetization transfers of this type occur as in-phase to anti-phase (or anti-phase to in-phase) transfers of weakly coupled systems. In protein NMR spectroscopy, the most common 2D experiments that feature these magnetization transfers are the HC-HSQC, and the HN-HSQC. In each of these experiments, magnetization is transferred from proton (denoted by H_x) to the heteronucleus (either ^{15}N or ^{13}C) during a pulse sequence block called INEPT (insensitive nucleus excitation by polarization transfer). Once magnetization is transferred such that the heteronucleus is anti-phase with respect to proton ($2H_zC_y$ or $2H_zN_y$), coherences evolve under the chemical shift of the heteronucleus during the incremented period t_1: $2H_zN_y \exp(-i\omega_H t_1)$. Following this so-called "encoding" period, magnetization is transferred back to proton $H_x \exp(-i\omega_N t_1)\exp(-i\omega_H t_2)$. 2D Fourier Transform gives a spectrum in which cross peaks appear at the chemical shifts of H and N; since the coherence transfer mechanism involves a weakly coupled system, a peak on a 2D HN/HC-HSQC is evidence of an H-N or H-C covalent bond. Indeed, the HN-HSQC is one of the most widely employed NMR experiments since each amino acid (except proline) gives a single signal (amide proton to its attached nitrogen). Moreover, chemical shifts are remarkably well dispersed along the N dimension, which makes the HN-HSQC a convenient "fingerprint" experiment that can either be used to probe for interacting partners (other proteins or small molecule) and concatenated to serve as a read-out for more complex experiments.

ii) Magnetization transfers through TOCSY-type transfers

In TOCSY (total correlation spectroscopy) experiments, a series of low power pulses (TOCSY pulse train) is applied to the sample. These highly engineered and optimized TOCSY pulse trains are referred to as isotropic mixing sequences because they effectively allow magnetization transfer to be modeled as evolution under the strong coupling Hamiltonian ($\mathcal{H}_{SC} \equiv \widehat{J_x} + \widehat{J_y} + \widehat{J_z} = 2\widehat{H_x C_x} + 2\widehat{H_y C_y} + 2\widehat{H_z C_z}$). The duration of the mixing sequence is set such that magnetization is passed in-phase to in-phase along the J-coupling network, leading to the term total correlation spectroscopy. Whereas in the case of HSQC-type experiments, magnetization is only transferred form one nucleus to other directly coupled (bonded) nuclei, in TOCSY-type experiments, magnetization can be transferred among all nuclei that exist in a coupling network. In the case of amino acid side chains, magnetization can be passed form one carbon in a side chain to all others. For example, In a 2D H(CC)H TOCSY experiment, several signals in the direct dimension will map to a single signal in the indirect dimension giving rise to so-called "TOCSY towers" that reflect all spins within a given coupling network. Whereas weak coupling transfers are directional, allowing for magnetization to pass from one nucleus directly to another, TOCSY transfers are diffuse along the side chain, such that magnetization is transferred to all other nuclei in the coupling network. TOCSY transfers are less sensitive than COSY-type transfers because optimal transfer times are typically longer than in the case for weak coupling transfers, but they provide superior information about the entire coupling network.

iii) Magnetization transfers through NOESY transfers

The NOESY experiment allows for magnetization transfer to be passed from one spin to another spin that is nearby in space. Whereas the previously discussed mechanisms of magnetization transfer occur through transferring coherence via J-coupling, NOESY type transfers occur through dipolar interactions. In NOESY transfers, the population excess of one spin can be passed to another. Dipole-dipole coupling allows for this population transfer while spins are of longitudinal spin order (aligned along +*z* or –*z*), and since they are facilitated by a dipolar mechanism, the intensity of transfer is inversely proportional to the spatial separation of the two spins:

$$NOE = \frac{1}{r^6}$$

Whereas the HSQC-type and TOCSY-type transfers reveal networks of spins bound to other spins, NOESY-type transfers reveal spatial proximity of spins.

3. NMR spectroscopy of biological macromolecules: assignment techniques, structure determination, and practical considerations

i) Motivation for development of NMR resonance assignment techniques

NMR spectroscopy is a powerful tool for functional and structural studies of proteins in solution and can provide atomic level information of structural, kinetic, thermodynamic, and dynamic parameters. This is because NMR signals can be determined precisely (sharp spectral lines) and because NMR parameters are exquisitely sensitive to subtle structural and dynamical changes. Indeed, the chemical shift responds to changes in the electronic environment, and nOe cross-peaks are affected by local structure (as are some ^{3}J coupling constants and residual dipolar couplings). Moreover, dynamical processes manifest in changes of signal linewidth and can be analyzed through a variety of relaxation experiments. No matter what the goal of the NMR experiment – structural studies, ligand binding studies, or dynamical studies – an essential step in any proper investigation requires the assignment of NMR signals, which entails the pairing of spectroscopic signals to nuclei of the polypeptide sequence. The appropriate combination of isotope labeling schemes and NMR pulse sequences that allows for NMR resonance assignment has been extensively reviewed[2–4]. For this introduction, the discussion is limited to large, structured proteins, for which many challenges must be overcome. Certain pulse sequences are highlighted because they are used specifically for large proteins. Their experimental implementations are discussed, and the manner in which they lead to resonance assignment is explicitly stated. The following sections are arranged according to sample labeling schemes, so as to showcase the pulse sequences that best exploit the properties of the respective samples. Rather than describing experiments with the density operator formalism, we have designed a graphical representation of pulse sequences that highlights common features. Occasionally, rough estimates of limiting molecular weights are provided, but the physicochemical properties of individual proteins alter spectral qualities greatly, so these numbers ought not be considered definite.

ii) General consideration for NMR spectroscopy of large molecular weight proteins & information requirements for structure determination

Larger proteins suffer from rapid relaxation (large R_2), which leads to sensitivity losses due to increased line broadening. Four major advances aimed at overcoming these issues have made NMR studies of large, monomeric proteins possible: improvements in hardware design, most notably the development of cryogenic probes to increase sensitivity by reducing noise, protein deuteration[5,6], TROSY[7,8], and non-uniform sampling (NUS)[9–11]. The TROSY method exploits the interference between two different relaxation mechanisms (most famously dipole-dipole & chemical shift anisotropy, CSA) that attenuates the relaxation of individual components of a multiplet. The slowly relaxing transition is selected with pulse sequence elements[8]. NUS consists of sparse sampling of data in indirect dimensions to drastically reduce experiment time.[12,13] Recent processing techniques have emerged to provide accurate spectra[9–11,14,15]. For larger proteins, the data acceleration is exploited to reach maximal spectral resolutions and to combat spectral overlap. NUS may also allow for recording more transients while maintaining resolution and, hence, improve sensitivity[16].

Determination of a protein NMR structure requires assigning resonances in NMR spectra. The following five steps summarize the assignment procedure: (i) Signals belonging to a single residue are grouped into a spin-system (e.g. denoted as {H, N, C^{α}}). (ii) Signals of sequential residues are identified, and the corresponding systems are linked to form sequential fragments. (iii) Side-chain signals are assigned to spin systems. (iv) Characteristic chemical shifts are used to assign systems to residues in the protein. (v) Finally, nOe cross-peaks between previously assigned protons are assigned to provide distance constraints. For large proteins, reduced sensitivity and increased spectral crowding impede all these steps; our text describes the spectroscopic techniques and sample labeling strategies that can be employed to overcome the challenges.

a) TROSY: transverse relaxation optimized spectroscopy

Wütrich and colleagues revolutionized biological NMR spectroscopy in the late 1990's with the advent of the TROSY principle: transverse relaxation optimized spectroscopy[17]. TROSY refers to an NMR experiment that consists of pulses and delays that specifically select the most slowly relaxing transition. This discussion is divided into two sections: theoretical basis of TROSY (spin physics and relaxation) and implementation of TROSY (basis of line selection pulse sequences).

Theoretical basis of TROSY

TROSY exploits differential relaxation mechanisms. For spin ½ nuclei, two mechanisms give rise to transverse relaxation through molecular tumbling: chemical shift anisotropy (CSA) and dipole-dipole (DD) interactions. The relaxation rate of a particular transition depends on the relaxation mechanisms affecting the nuclei participating in the transition; in the case of cross-correlated relaxation two different relaxation mechanisms affect the relaxation rate of the transition (for instance, CSA and DD). Fortunately, a complex master equation has been developed that allows prediction of the effects of cross-correlated relaxation on transitions. For an H-N system, the effects of cross-correlated relaxation on the H^+N^β transition lead to a decrease in the relaxation (TROSY). In TROSY sequences, pulses are arranged such that this transition is detected, which may be accomplished using single transition to single transition polarization transfer (ST2-PT)[17] or spin state selective excitation (S3E) schemes[18].

Practical Implementation of TROSY: ST2-PT & S3E

As mentioned above, TROSY involves a selection of a single transition from a multiple spin system (such as, the H-N spin system above); consequently, single transition operators (i.e. H^+N^β) were used to understand the TROSY principle. In order to understand how the pulse sequences facilitate selection of this line, I will be changing to a Cartesian product operator basis, which allows for understanding how pulses and delays affect the NMR signals (occasionally, polarization operators will be used, e.g. I_xS^β). The first example of a TROSY experiment is referred to as the single transition to single transition polarization transfer experiment, or ST2-PT[17]. The ST2-PT experiment is shown in Figure 1.1 with product operators at each step along the pulse sequence. The ST2-PT block converts the product operators ($2I_zS_y$ - S_y) into ($2I_xS_z + I_x$) and ($2I_xS_z$ - S_x) into ($I_y + I_yS_z$) during two concatenated INEPT periods; recast into polarization operators, the TROSY effect is more obvious as $I^\beta S_y$ → $-I_xS^\beta$ and $I^\alpha S_x$ → $-I_yS^\alpha$.

The experiment and signal algebra in *Figure 1.1* is referred to as "phase-selected TROSY" experiments since changing the phases of the specified pulses and receivers results in selection of the TROSY-line. Later, pulse field gradients were used to select for the TROSY line[19–22]. By minimizing transverse relaxation losses during transfer of ^{15}N to $^{1}H^{N}$, Yang and Kay improved sensitivity of gradient selected TROSY[23,24]. Finally, Neitlispach realized that adjusting the phases of the 90° pulses in the sensitivity improved, gradient HN-TROSY scheme (of Yang and Kay) could reduce signal artifacts (specifically of the anti-TROSY lines)[25]. The HN-TROSY sequences in this thesis use either the standard ST2-PT or Nietlispach-style schemes. An alternative strategy to select a single transition in one scan uses S3E-type experiments for (spin-state selective excitation, S3E)[26,27]. In these experiments, 45° pulses are applied to longitudinal systems. After evolution under J_{HN} for $\frac{1}{4J_{NH}}$, the single transition operators are orthogonal to each other in the transverse plane. Application of a hard 90° pulse on ^{1}H with either phase *x* or *y* can selectively capture either I_yS^{β} or I_xS^{α}.

b) Protein perdeuteration & partial deuteration

Since the early days of protein NMR spectroscopy, the enormous benefit of replacing ^{1}H protons with ^{2}H deuterons has been exploited. Since the dipolar contribution to a spin's relaxation rate scales with the square of the gyromagnetic ratios of the nuclei involved in the dipolar interaction[28], replacement of aliphatic and aromatic ^{1}H with ^{2}H $((\gamma^{1}H)/(\gamma^{2}H) \sim 6.5)$ vastly improves relaxation of carbon coherences and to a lesser extent, amide proton and nitrogen coherences. In 1968, Crespi et al. grew bacteria in ^2H_2O and analyzed the chemical shifts of the few remaining, non-exchangeable nuclei as observables for ligand/protein binding or structural rearrangements[29,30]. Meanwhile, Jardetzky and coworkers expressed *Staphylococcal* nuclease in a perdeuterated background and exogenously supplied otherwise protonated amino acids (tyrosine, methionine, or tryptophan) for one-dimensional NMR analysis[6]. Thus, deuteration had been an accepted manner to simplify one-dimensional spectra. After two-dimensional NMR spectroscopy became increasingly common, random fractional deuteration was used as a tool to decrease linewidths and obtain optimal relaxation properties. Furthermore, dilution of ^1H with ^2H minimized ^1H-^1H spin diffusion in NOESY spectra. In the early 1990's, random fractional deuteration of proteins became more common to increase the sensitivity of heteronuclear experiments. Grzesiek used random fractional deuteration to study the 20 kDa protein calcineurin[31]; LeMaster used the same strategy to assign the 12 kDa thioredoxin[32]. Muhandiram and Kay used random fractional deuteration to study ^{13}C relaxation properties of a SH2 domain in complex with a peptide (~14 kDa)[33]. Around this time, the Kay lab had already developed pulse sequences on highly deuterated proteins, some of which were published using data collected on partially deuterated proteins[34,35]. Nietlispach attempted to find the right balance of ^1H_2O and ^2H_2O and for structure determination of larger proteins with NMR spectroscopy using specific experiments[36]; however, his estimates were published before the widespread application of the TROSY effect and included experiments that are generally too insensitive for larger proteins (4D HBCB/HACA(CO)NNH, 4D HBCB/HACANNH).

Following some of the pioneering work of Venters and Farmer[37], in which protein was expressed in 100% $^{2}H_2O$, $^{15}NH_4Cl$, and $^{1}H^{13}C$-sodium acetate, which allowed for perdeuteration of the protein backbone with some reprotonation of various nuclei throughout the protein, Rosen, Gardner, and Kay later had the notion of using $^{1}H^{13}C$-sodium pyruvate in an otherwise $^{2}H^{13}C^{15}N$ background, which allowed for labeling of most methyl groups[38-28]. Unfortunately, several amino acids are not labeled in the pyruvate strategy, notably the δ1 of Ile methyl group, which is especially important because it is well resolved in $H^{methyl}C^{methyl}$ HC-HSQC or HC-HMQC spectrum, owing to its distinctive $^{13}C^{\delta1}$ chemical shift. To this end, Gardner and Kay began using ^{13}C α-ketobutyrate to protonate δ1 of Ile[40]. Finally, Goto and Kay began using [3,3-$^{2}H_2$,^{13}C] α-ketobutyrate and [3-^{2}H, ^{13}C] α-ketoisovalerate, which gives rise to (^{1}H-δ1 methyl)-isoleucine, (^{1}H-γ methyl)-valine, and (^{1}H-δ methyl)-leucine (*Figure 1.2)*[41].

c) NUS: non-uniform sampling & reconstruction techniques

Non-uniform sampling or NUS is a data collection strategy by which only a subset of points is acquired in indirect dimensions[42]. Traditionally, the time-domain of indirect nuclei are acquired by indirectly sampling the dimension of interest with spacing between points (i.e. the "dwell time") equal to the inverse of the spectral width of the nucleus of interest $\Delta t_i = 1 / SW_i$. The resolution of these indirect dimensions is $N_i * \Delta t_i$, where N_i is the number of acquired data in the indirect dimension. In multi-dimensional NMR experiments, indirect dimensions are sampled independently from each other; consequently, experimental time increases geometrically with dimensionality and resolution. In the cases of some three-dimensional NMR spectra and nearly all four-dimensional NMR spectra, the experimental acquisition time is prohibitive for optimal resolution. In NUS, only a subset of these indirect dimension points is acquired, which allows otherwise prohibitively long NMR experiments to be recorded within practical limitations of instrument availability and specimen lifetime. If N_{max} refers to the number of indirect dimension points desired, and N_s refers to the subset of indirect dimensions points sampled, then we can define a sampling factor $s = N_s / N_{max}$, which would result in acceleration of data collection by a factor of s^{-1}. The choice of sampling factor itself depends upon the size of N_{max}. When N_{max} corresponds to indirect dimension times that approach the signal's natural linewidth (the point beyond which there are diminishing returns for sampling the signal, $t_{1naturalmax} = 1.2 / R_2$)[43], then the sampling factor can be small because NMR signals are sparse within N_{max}. In contrast, when N_{max} corresponds to sampling lower than the signal's natural linewidth, sampling factors need to be larger because signals are more concentrated within N_{max}[44].

The choice of which points to sample is referred to as the "sampling schedule," and the generation of the best sampling schedule has been the subject of significant work within NMR non-linear sampling field. For all work in this thesis, the PoissonGap[45,46] sample schedule generator has been used. The PoissonGap software allows the user to estimate the relaxation rate of various coherences as a function of magnetic field and molecular weight; these estimates are useful for determining best value of N_{max} if NUS is being used to achieve the natural linewidth of the signals. It is fairly well accepted that data points should be randomly sampled from the indirect dimension time domains[47]; however, PoissonGap[46] imposes a few restrictions on the randomization of sampled points: large spaces (gaps) between sampling points is prevented and gaps are minimized at the beginning and end of the indirectly sampled time-domains.

NUS data do not give rise to a faithful spectrum through a simple series of direct Fourier Transforms, apodizations, and phase shifts, as is routinely the case for uniformly sampled data. When sparsely sampled data are subjected to Fourier transform, indirect dimension points that are not sampled are effectively given zero values; thus the data matrix contains only s points of data, and missing points are filled in with zeroes. When this so-called "zero augmented" data matrix is subjected to Fourier Transform, the resulting frequency domain contains many artifacts and features poor signal-noise ratio. Indeed, the Fourier Transform does not handle discontinuities well, and the resultant "spectrum" is a result of the uniformly sampled spectrum convolved with the point-spread function (that is, the Fourier Transform of the sampling schedule). Since simple, direct Fourier Transforms do not yield a quality spectrum, other mathematical and signal processing routines have been developed to reconstruct spectra—sometimes these are referred to as "non-Fourier methods" even though they all rely in some capacity on use of the Fourier Transform. Several of these reconstruction techniques are readily available and have been extensively reviewed[10,11,48–50]. This thesis contains NUS data that were all reconstructed using the istHMS (iterative soft thresholding from Harvard Medical

School)[46,50]. The major criteria for selection of istHMS were the quality of the reconstructed spectra, the superior speed in reconstruction time, and the lack of reliance upon a particular type of sampling schedule (as in radial sampling[51], for example). istHMS begins by running the Fast Fourier Transform (FFT) algorithm on the sparsely acquired time domain data, which produces a spectrum containing a large amount of artifacts (because of convolution with the point spread function). At this point, the most intense frequency domain signals are cut-off and stored elsewhere in computer memory. The truncated spectrum then undergoes inverse FFT and the process is iterated until termination criteria have been reached (usually a user-defined parameter – for all of the spectra in this thesis work, 400 iterations was used as the termination criterion).

d) Covariance NMR spectroscopy

Because of disastrous signal overlap in large proteins, peak-picking (determining the chemical shift of a nucleus based on its signal maximum) is often subject to human error. Indeed, it can be extremely difficult to properly identify signal maxima of severely overlapped signals. To circumvent human error in peak-picking, we have developed some covariance NMR based protocols that remove the human component of peak-picking, making it easier to properly pair signals of successive backbone resonances. These techniques are more fully described in Chapter 3.

e) Time-shared Principle

Conventionally, INEPT magnetization transfers in NMR experiments occur from one set of nuclei to another (for example, $^{1}H \rightarrow {}^{15}N$). A 90° pulse puts magnetization into the transverse plane and after a delay period of $1/4J_{HN}$, 180° are applied to both nuclei. Another delay period of $1/4J_{HN}$ follows, and finally a 90° pulse on the first nuclei completes the INEPT; in the product operator formalism, this block effectively converts H_z to $2H_zN_z$. Using the time shared principle[52–54], a third 180° pulse is applied to a third channel, say, to the ^{13}C channel, during a period that is ($¼J_{HN}$ – $¼J_{HC}$). Invocation of another channel in this INEPT allows for two simultaneous magnetization transfers: ($H_z \rightarrow 2H_zC_z$ and $H_z \rightarrow 2H_zN_z$). In essence, two different coherence pathways can be initiated from a single starting block. The time-shared principle is used in the HN-TROSY/HC-methyl-TROSY-HMQC-NOESY experiment that is described in detail in Chapter 4.

4. A survey of NMR experiments used for assignment & structure calculations of high molecular weight proteins

Having described the important building blocks and key aspects of NMR experiments on biological macromolecules, the following sections describe various isotopic labeling strategies of NMR samples and the NMR experiments that commonly accompany them for resonance assignment and structure determination.

i) Preliminary samples: ^{1}H , ^{15}N samples

A ^{1}H-^{15}N-labeled sample (sometimes partially deuterated) is often produced at the onset of a project to verify that the protein is amenable for NMR investigations and to optimize measurement conditions. For example, HN-TROSY[55] reports on the state of the protein (folded, disordered, aggregated, degraded), helps determine the spectral width in the ^{15}N dimension, and allows for inspection of protein dynamics. Recording an H/N correlation map in D_2O and monitoring signal intensities leads to identification of slowly exchanging protons. Protons still detected after extensive times (days/weeks) will be absent in spectra of deuterated samples typically used for NMR measurements in larger proteins (*Chapter 3 Section 3*). A useful starting point is the ^{15}N-TRACT, which evaluates the correlation time (a sensitive reporter of aggregation) by estimating the relaxation rates of the TROSY and anti-TROSY transitions[56]. NOESY-HN-TROSY-HSQC can be recorded with a ^{1}H^{15}N sample for proteins up to 40 kDa, albeit with mixing times shorter than for smaller proteins to minimize spin diffusion. Such a spectrum will provide valuable short distance constraints, for example between amide and alpha protons or between nearby amide protons. The latter may be critical for sequential assignment of backbone resonances when unfolding/refolding of the protein cannot be performed to rescue slowly exchanging protons in deuterated samples (*Section 4.v*). While only a small subset of constraints might be detected and assignable, they may nevertheless help refine structures, as exploited for a 37 kDa protein[57]. Lastly, for proteins up to ~45 kDa, HH-TOCSY-HN-TROSY-HSQC[8] (*Figure 1.4 X*) can supply intra residue (H,H) correlations. In particular, H^{α} assignment provides access to H^{N}-H^{α} distance restraints, which help lower the root-mean-squared-standard-deviation of structural ensembles, especially for β-sheets. MQ-HACACO provides an alternative means to assign H^{α} resonances (*Section 4.v,* below).

ii) The workhorse sample for backbone resonance assignment: $^{2}H^{15}N^{13}C$

For larger proteins, backbone assignment is usually performed using a ^{2}H-^{13}C-^{15}N sample. Deuteration replaces all non-exchangeable protons with deuterons and improves both sensitivity and resolution in NMR spectra, thus rescuing NMR applications for larger proteins. Sequential correlations are traditionally obtained with a suite of triple resonance experiments that correlate $\{H^{N}, N\}$ systems with either C^{α}, CO, or C^{β} carbons, for the same *and* preceding residues. Thus, correlations between amide groups and backbone carbons are provided directly, while links between sequential $\{H^{N}, N\}$ systems result indirectly from identifying shared correlations to backbone carbons, e.g. $(H^{N}_{i}, N_{i}, C^{\alpha}_{i})$ and $(H^{N}_{i+1}, N_{i+1}, C^{\alpha}_{i})$. HNCA (*Figure 1.3 I*), HN(CA)CO (*Figure 1.3 III*), and HN(CA)CB (*Figure 1.3 V*) provide both sequential (i-1) and intra-residue (i) correlations between $\{H^{N},N\}$ systems and C^{α}, CO, and C^{β} carbons respectively. In practice, the sequential correlations may not be observed, as noted for the 81 kDa malate synthase G (MSG)[58]. HN(CO)CA (*Figure 1.3 IV*) and HNCO (*Figure 1.3 II)* provide selectively sequential correlations $(H^{N}_{i}, N_{i}, C^{\alpha}_{i-1})$ and $(H^{N}_{i}, N_{i}, CO_{i-1})$. Together, these experiments provide systems of the form $\{H^{N}_{i}, N_{i},$

$C^{\alpha}{}_{i}$, $C^{\alpha}{}_{i-1}$, CO_i, CO_{i-1}, $C^{\beta}{}_{i}$, $C^{\beta}{}_{i-1}$} and link sequential systems by identifying a triplet of carbon frequencies (C^{α}, CO, and C^{β}) that is common to two such systems.

While this strategy is used effectively for proteins up to 25 kDa, it often fails for larger proteins because (near) degenerate frequencies abound and because limited sensitivity results in absent NMR signals and hence incomplete spin systems. Thus, special considerations must be made for data acquisition, and additional experiments may be required. Implementing NUS to reach the maximal resolution in every dimension[59] is vital in preventing apparent degeneracies arising from limited resolution. All magnetization transfers should be optimized experimentally to maximize signal intensity. Usually, this strategy amounts to a reduction in the transfer duration to lessen relaxation. However, in HN(CA)CB, the transfer from C^{α} to C^{β} is optimized for C^{β} signal, and C^{α} is sacrificed, unlike in small proteins. The more sensitive HNCA and HN(CO)CA provide correlations to C^{α}. While HN(CA)CB is the most insensitive sequence in this set, C^{β} correlations are critical for (at least) three reasons. First, the set of inter- and intra-residue correlations provides an additional level of discrimination when building sequential links, and therefore, they improve the quality of sequential correlation maps (see *Chapter 3.2*). Second, many amino acids display characteristic C^{β} chemical shifts that allow for identification of the amino acid type. Finally, C^{β} chemical shifts can be used along with C^{α} and CO shifts to predict secondary structure and backbone torsion angles[60], which can be used during structure determination. For these reasons it is customary to monitor the sensitivity of a 1D H^N trace of HN(CA)CB to assess the feasibility of backbone resonance assignment.

iii) Protonated methyls in a deuterated background I: U-^{2}H-^{15}N-^{13}C ILV Me-^{1}H sample for methyl resonance assignment

In large proteins, methyl groups are the main source of side-chain distance constraints, and targeted strategies are used to assign their resonances. The assignment of methyl resonances is typically carried out using a sample that is ^{1}H-^{13}C at specific methyls against an otherwise ^{2}H-^{13}C-^{15}N background. This selective protonation reintroduces key probes for side-chain assignment while maintaining most of the favorable relaxation properties of a perdeuterated sample. Protocols have been established for the selective protonation of valine (Val)-γ and leucine (Leu)-δ[61], isoleucine (Ile)-δ1[61], Ile-γ2[62], alanine (Ala)-β[62], threonine (Thr)-γ2[63], and methionine (Met)-ε methyl groups[64,65]. Notably, the methyl groups of each amino acid can be labeled independently (except Val and Leu, which use a common precursor), providing flexibility in designing labeling schemes that simultaneously maximize observed correlations and minimize spectral overlap.

To assign the methyl groups, one seeks to establish correlations between methyl protons and previously assigned resonances of backbone nuclei. H(CCCO)NH and (H)C(CCO)NH (*Figure 1.4 I, II*) provide correlations between either the proton or carbon methyl resonances, respectively, and assigned {H, N} systems[66]. Although, these experiments are insensitive due to complex spin manipulations and because branching in Ile, Leu, and Val side-chains directs some magnetization away from the desired C^{α} during the TOCSY period[67], we will show that NUS can rescue sensitivity losses (*Chapter 3 Section 6*). Another approach from Tugarinov and Kay involved developing a set of experiments that employs a series of COSY transfers to relay magnetization from methyl groups to backbone nuclei[68]. HMCM(CGCBCA)CO (*Figure 1.4 VII*) uses INEPT and COSY transfers to supply correlations of the form (H^{M}_{i}, C^{M}_{i}, CO_{i}); thus, they correlate unassigned methyl groups with assigned carbonyl carbons. The equivalent experiment for Val is achieved by removing one of the COSY transfers (*Figure 1.4 VII* without the blocks enclosed in dashed lines). HMCM[CG]CBCA (*Figure 1.4 VIII*) uses a similar set of COSY transfers to provide (H^{M}_{i}, C^{M}_{i}, C^{α}_{i}) and (H^{M}_{i}, C^{M}_{i}, C^{β}_{i}) thus correlating methyls and assigned C^{α} and C^{β} carbons. These correlations also alternate in sign along the side-chain, such that for Ile and Leu, correlations to C^{α} are positive and those to C^{β} negative, whereas these signs are inverted for Val. Thus Val correlations, which occupy the same spectral region as Leu, can be identified unambiguously.

In summary, the combination of these experiments provides correlations between the methyl resonances of Leu, Val, and Ile (δ1 only) and previously assigned CO, C^{α}, and C^{β} resonances; therefore they provide the links necessary for assigning methyl resonances. Additionally, these experiments may complete the assignment of previously undetected or ambiguous C^{α}, C^{β}, or CO resonances, allowing one to extend backbone resonance assignment or confirm/correct ambiguous assignments.

Such experiments can be recorded either on samples where only one methyl of each Val and Leu is labeled or on samples where both methyls are labeled. Although the latter samples are cheaper, they will produce data of reduced signal to noise and with more artifacts.

The methyl groups of Ala, Ile-γ2, and Thr can be assigned similarly. For Ala, 4D-ALA-HMCBCACO[69] (*Figure 1.4 VII* without the blocks enclosed in solid boxes, chemical shift encoding is designated with white text) provides correlations of the form (H^{M}_{i}, C^{M}_{i}, C^{α}_{i}, CO_{i}) and directly links Ala methyl proton and carbon resonances to two backbone nuclei. Alternatively, if both Ala and Ile-γ2 are labeled, four 3D experiments can be used to simultaneously assign methyls of both amino acid types[70]. This set of experiments correlates methyl resonances with C^{α} and CO. Finally, Thr-γ2-selective 3D HMCMCBCA correlates Thr methyls with both C^{α} and C^{β}[71]. Met-ε methyl groups are assigned based on nOe cross-peaks with previously assigned methyls. Such a strategy must be performed with help from a structural model. Alternatively, one can mutate individual Met residues and assign these to resonances that disappear from an HC-HSQC.

iv) Protonated methyls in a deuterated background II: U-^{2}H-^{15}N-^{12}C ILV Me-^{1}H^{13}C sample for nOe measurement

Measurement of nOes used for structure determination is best performed on samples in which selected methyl groups are ^{1}H-^{13}C while the rest of the carbons are ^{2}H-^{12}C.

(*Chapter 4* provides an alterative to this strategy.) This scheme minimizes relaxation and prevents evolution under scalar couplings between methyl and other side-chain carbons and also between amide nitrogens and backbone carbons. As described in *Section 4.iii*, most methyl-bearing amino acids can be labeled independently of one another, allowing one to choose the combination of amino acids to label in order to minimize spectral overlap. Due to the scarcity of protons in such samples, it is necessary to maximize the number of detected nOes that will lead to constraints. One is obliged to accurately assign nOes and maintain enough sensitivity to accurately integrate signal intensities. *Correct* assignment is crucial, as misassignment of only a few nOes can lead to poor or erroneous convergence during structure determination.

Because of spectral crowding in H/N correlation maps and limited chemical shift dispersion of methyl proton resonances, unambiguous assignment of nOes is best done using 4D experiments. A 4D-NOESY provides a two-dimensional array of H/N or H/C correlation maps, in which the array coordinates also represent H/N or H/C correlation maps. Each 2D plane associated with each ($H^N{}_i$, N_i) or ($H^C{}_i$, C_i) coordinate in the array is an HSQC containing cross-peaks only for protons H_j with a dipolar interaction to H_i. A simple comparison of these 2D planes with an assigned 2D HN- or HC-HSQC provides rapid and efficient assignment of nOes.

The time-shared-4D-HC-HSQC/HN-HSQC-NOESY-HC-PEP-HSQC/HN-TROSY (TS-4D) experiment allows for the assignment of *all* amide and methyl nOes in a methyl-only labeled sample from a single 4D experiment [72]. This experi=-0 nt uses the time-shared principle both before and after the mixing period. Therefore, all four possible pathways ($H^N \rightarrow H^N$, $H^N \rightarrow H^M$, $H^M \rightarrow H^N$, and $H^M \rightarrow H^M$) are recorded. Post-acquisition processing deconvolutes the four pathways and provides all four possible 4D experiments. For higher sensitivity in the $H^M \rightarrow H^M$ 4D, TS-4D can be supplemented with 4D-HMQC-NOESY-HMQC [73] that employs methyl-TROSY (HMQC) both before and after the mixing period, in contrast to TS-4D. These experiments can also be recorded using NUS, allowing one to maximize resolution without resorting to prohibitive acquisition time[74].

The sensitivity of a 4D experiment may be too low for reliable integration of all signals, so 3D TS-(HN-TROSY/HC-HSQC)-NOESY may be used as a complement for weaker signals[75]. This experiment uses the time-shared principle to provide both a 3D-HC-HSQC-NOESY and 3D-HN-TROSY-NOESY from a single experiment. Critically, the nOe cross-peaks appear in the detected dimension, i.e. with maximum resolution. Such a resolution minimizes overlap of resonances and improves the accuracy of signal integration. (See *Chapter 4*) In summary, the use of multiple samples with different labeling schemes, assignment of nOes using 4D experiments, and integration using a sensitive 3D experiment allows for the structure determination of large proteins.

v) ^{1}H-^{13}C-^{15}N sample for rescue of non-exchangeable amide protons

In larger proteins, many amide groups participate in hydrogen bonds, which reduces their ability to undergo solvent exchange, and in deuterated samples these amide groups will escape detection. Refolding the protein from its denatured state is an efficient way to facilitate amide deuteron exchange (*Chapter 3 Section 4*); however, many proteins do not tolerate such a treatment. In such cases, it may be necessary to use ^{1}H-^{13}C-^{15}N-labeled samples to complete resonance assignment, despite their poor spectral properties. Of the backbone experiments described hitherto only HNCO and HNCA are likely to deliver enough sensitivity. All experiments with magnetization transfers involving aliphatic carbon coherences will suffer from disastrous dipolar relaxation. Thus, alternative strategies must be employed to identify sequential residues. In particular, ^{13}C-detected (^{13}C-d) experiments for backbone and side-chain assignment have emerged thanks to recent advances in cryoprobe technology. Despite a reduction in sensitivity by a factor of $((\gamma^{1}\mathrm{H})/(\gamma^{13}\mathrm{C}))^{3/2} \sim 8$,[76] ^{13}C-d experiments allow for reduction of the number of magnetization transfers needed to generate correlation maps (by about half) and, therefore, partially compensate for relaxation losses, which can be severe for protonated samples. Using NUS to increase the number of transients applied while maintaining conventional resolution can compensate for the lower sensitivity of ^{13}C-d experiments.

In ^{13}C-d experiments, special methods must be employed to account for evolution under homonuclear couplings during detection. Specifically, multiplets are deconvoluted during processing[77,78] or spin-state selection is used to obtain single signals[79–81]. Since refocusing ^{13}C-^{13}C couplings requires long transfer times, oftentimes one detects coherences anti-phase with respect to coupled carbons[82], and spectra must be manipulated during processing to provide a single signal[83,84].

Sequential backbone resonance assignment may be achieved with a combination of ^1H-detected (^1H-d) and ^{13}C-d experiments. The sensitive HNCO and HNCA are analyzed together with MQ-HACACO (*Figure 1.3 XI*) that provides the correlation ($H^{\alpha}_{i,}$ C^{α}_{i}, CO_i)[85,86]. MQ-HACACO takes advantage of MQ spectroscopy (quenching much of the proton-carbon dipolar relaxation), concatenation of two CT periods (for H^{α} and C^{α}), and rapid pulse recycling with optimized H^{α} magnetization recovery[87] (so-called "H-flip" experiments)[88] to improve sensitivity. Relating {H^{N}_{i}, N_i, CO_{i-1}} spin systems from HNCO, with {$H^{\alpha}_{i-1,}$ C^{α}_{i-1}, CO_{i-1}} from MQ-HACACO and {H^{N}_{i-1}, N_{i-1}, C^{α}_{i-1}} from HNCA allows for sequential resonance assignment. The method was applied to a 44 kDa 35% deuterated sample[86] and a 37 kDa 100% protonated sample[57].

Side-chain ^1H and ^{13}C assignments in ^1H-^{13}C-^{15}N samples may be obtained using either NOESY- or TOCSY-mediated carbon to carbon magnetization transfers. Zuiderweg proposed using ^{13}C-^{13}C short-range nOes in 3D-(H)CCH-NOESY or 4D-HCCH-NOESY (*Figure 1.3 IX*) to obtain correlations between aliphatic groups[89]. Protons are correlated to carbons with INEPTs, but adjacent carbons are correlated via nOe's rather than scalar couplings. Thus, the density operator is subject to small longitudinal rather than large transverse relaxation rates. Alternatively, CT and RT ^{13}C-excited (^{13}C-e)/^{13}C-d-TOCSY (*Figure 1.4 V*) have been designed to minimize magnetization transfers and associated relaxation losses[90]. Although the CT implementation of this experiment is longer than its brother RT experiment, it allows for both a gain in resolution and for the identification of carbon substitution multiplicity from the sign of the corresponding cross peak (as in DEPT[91]). Comparing ^{13}C strips of 2D ^{13}C CT-TOCSY to assigned signals in HNCA, HN(CA)CB and MQ-HACACO allows for carbon side-chain resonance assignment. ^{13}C - and ^{15}N -edited NOESY are subsequently used to assign ^{1}H side-chain resonances from assigned carbon resonances.

vi) Partially deuterated samples: cost effective sources of nOe cross-peaks and favorable relaxation properties

As described in *Section 4.iii and Section 4.iv*, to overcome relaxation and maintain spectroscopic probes, special efforts must be made to reintroduce protons into otherwise perdeuterated samples. Such endeavors come at significant monetary cost, and in favorable cases, a fractionally deuterated sample may be sufficient to ameliorate relaxation properties without sacrificing crucial ^{1}H-^{1}H nOes, making assignment tractable[32]. Clearly, the fraction ^{2}H/^{1}H cannot be simultaneously optimized for all experiments traditionally used for backbone and side-chain resonance assignment. The discussion that follows may apply to samples that vary from 35% to 70% deuteration.

Many of the backbone experiments described in *Section 4.i* can be recorded on a fractionally deuterated sample, albeit with modifications. We advise experimental comparison of the sensitivity of experiments where TROSY-line selection is achieved with so-called sensitivity-enhanced-like schemes (which employs in part longitudinal magnetization of slower relaxation[25]) to those containing ST2-PT elements with N→CA/CO concatenation (shortening the sequence by 5.5. ms)[92]. The relative efficiency of these two schemes is likely to depend on the level of deuteration. Evolution under both ^{1}H-^{13}C and ^{2}H-^{13}C scalar couplings must be refocused. While the latter can be achieved with composite-pulse-decoupling as for ^{2}H-^{13}C-^{15}N samples, proton inversion pulses are used to refocus evolution under $^{1}J_{CH}$ without affecting TROSY[93]. Note that targeting the detection of both protonated and deuterated isotopologues (same chemical composition, different isotopes) imposes limits on the resolution, such that the two signals will not be resolved. Consequently NUS should not be used to achieve maximal resolution; however, it can be exploited to increase the sensitivity of these sequences by increasing the number of transients and maintaining appropriate resolution[16] (*Chapter 3 Section 1*).

As with protonated samples, both NOESY- and TOCSY-mediated transfers facilitate side-chain resonance assignment of fractionally deuterated samples. 2D ^{13}C-^{13}C NOESY (*Figure 1.3 VI*) with long mixing times (300-800 ms) can be used to observe one-, two- and sometimes three-bond connectivities through spin diffusion, which allows for linkage of C^{β} and C^{γ} to CO[94]. Because the maxima in spin diffusion/nOe buildup curves occur at shorter mixing times for increased correlation times these experiments may be useful for larger proteins [95]. The HCCH-TOCSY experiment traditionally used for protonated small proteins displays poor sensitivity in fractionally deuterated samples because of unfavorable isotopic distribution. Rather, conventional H(CCCO)NH and (H)CC(CO)NH (*Figure 1.4 I-II*) are supplemented with ^{13}C-d or ^{13}C-e experiments. HCC-TOCSY (*Figure 1.4 IV*)[85] is the ^{13}C-d equivalent of the well-known HCCH-TOCSY. Pervushin modified this sequence such that ^{13}C chemical shifts are encoded during MQC evolution periods (MQ-HCC-TOCSY) to improve sensitivity[96]. Using antiphase coherences before the TOCSY mixing sequence further improves sensitivity but requires a laborious fitting procedure to deconvolute detected multiplets. In place of a ^{1}H-excited (^{1}H-e)/^{13}C-d HCC-TOCSY (~8x less sensitive than ^{1}H-e/^{1}H-d), Mulder et al. developed a ^{13}C-e/^{1}H-d CCH-TOCSY (~4x less sensitive than ^{1}H-e/d). CCH-TOCSY (*Figure 1.4 V*) features sensitivity enhancement in two dimensions and selects for CHD_2 isotopologues. It provided assignment for ~90% of the methyl groups of a 34 kDa protein but required 1 mM concentration and ~60h data acquisition. NUS may be employed to lessen acquisition time or sample concentration requirements.

Partially deuterated samples have a distribution of isotopologues, and it is often best to select signals of a particular isotopologue to minimize spectral crowding. This effect is particularly problematic for methyl groups that feature four isotopologues CH_3, CD_3, CH_2D, and CHD_2 (the latter two usually represent more than 80% of the distribution), so it is critical to select only a single isotopologue[38,97]. Under most circumstances CHD_2 resonances are sought because of their relaxation properties[33]. Adjusting the delay τ in a refocused INEPT so that $2\pi^1J_{CH}\tau$ equals the magic angle (54.7°) leads to optimal CH_3 suppression, albeit at the cost of reduced sensitivity[98].

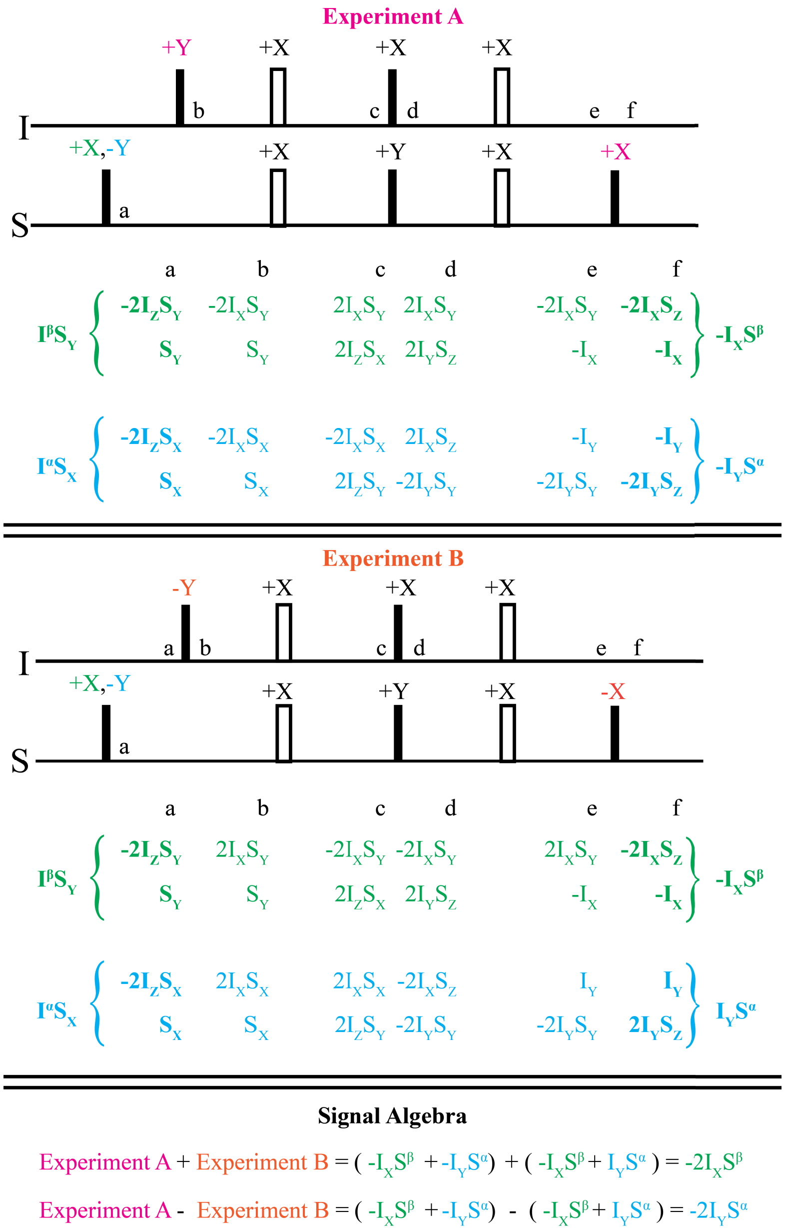


Figure 1.1 ST2-PT pulse sequence element allows for selection of a single line. The density matrix operators are expressed in terms of polarization operators outside the braces; inside the braces, density matrix operators are expressed in the conventional Cartesian product basis. ST2-PT TROSY[17] is achieved by running two experiments (**A** and **B**) separately; for both **A** and **B**, $\mathbf{I^{\beta}S_y}$ and $\mathbf{I^{\alpha}S_x}$ are selected by choice of the phase of the first 90˚ pulse on the S channel (**+x** or **–y**). In experiment

B, two phases are changed by 180° (**magenta** and **orange**), which allows for change of sign of $\mathbf{I_yS}^{\alpha}$. The resultant signals can be added or subtracted to select only the $\mathbf{I_xS}^{\beta}$ or $\mathbf{I_yS}^{\alpha}$ transitions in the final spectrum. Note that we detect I_-.

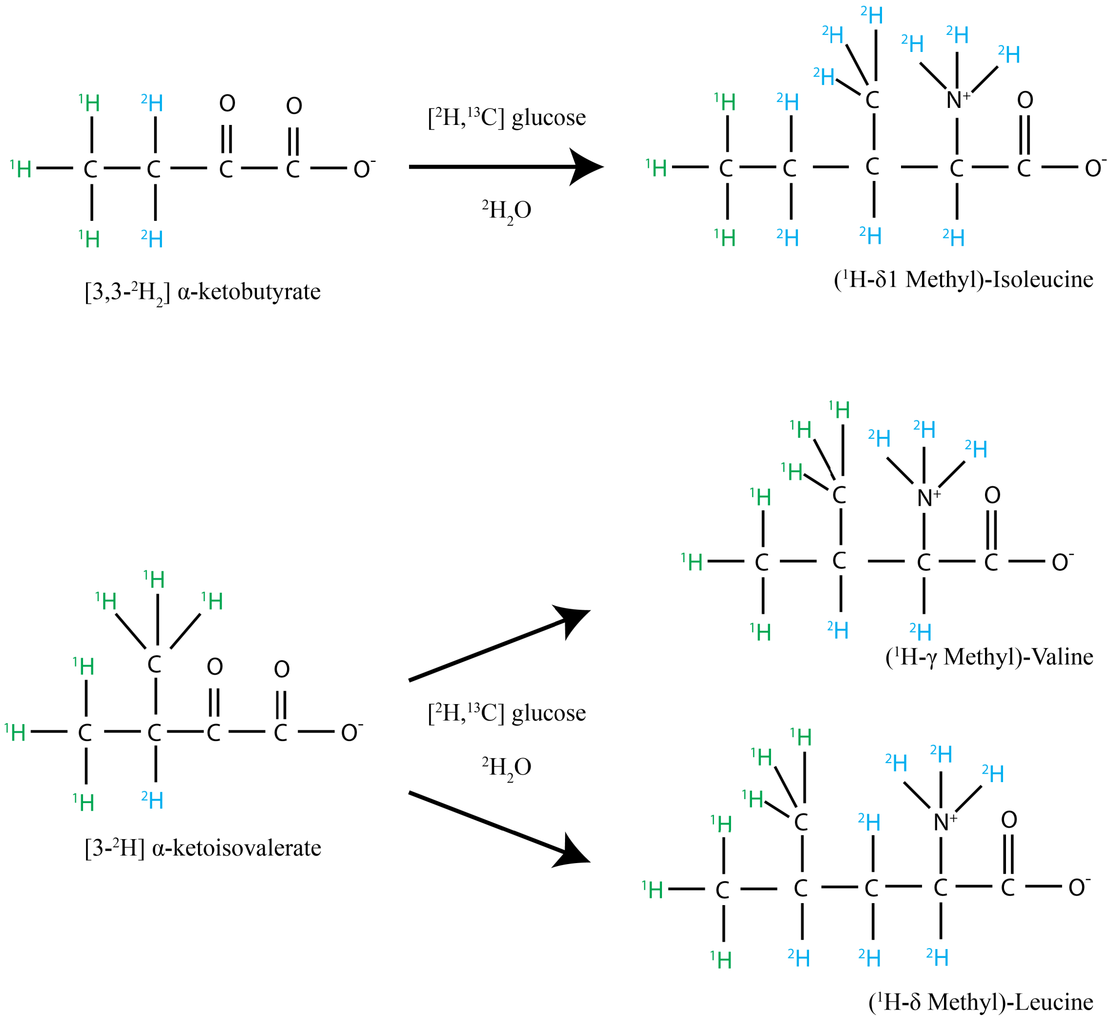


Figure 1.2 Commonly used precursors to accomplish ^{1}H Methyl ILV labeling in a perdeuterated background. The labeling approach of Goto et al.[41] is represented. On the left the precursors [3,3-$^2H_2$] ^{13}C α-ketobutyrate and [3-$^2H_2$] ^{13}C α-ketoisovalerate are shown, with ^1H in green and ^2H in blue. On the right, the specific labeling of the resultant Isoleucine, Valine, and Leucine is shown.

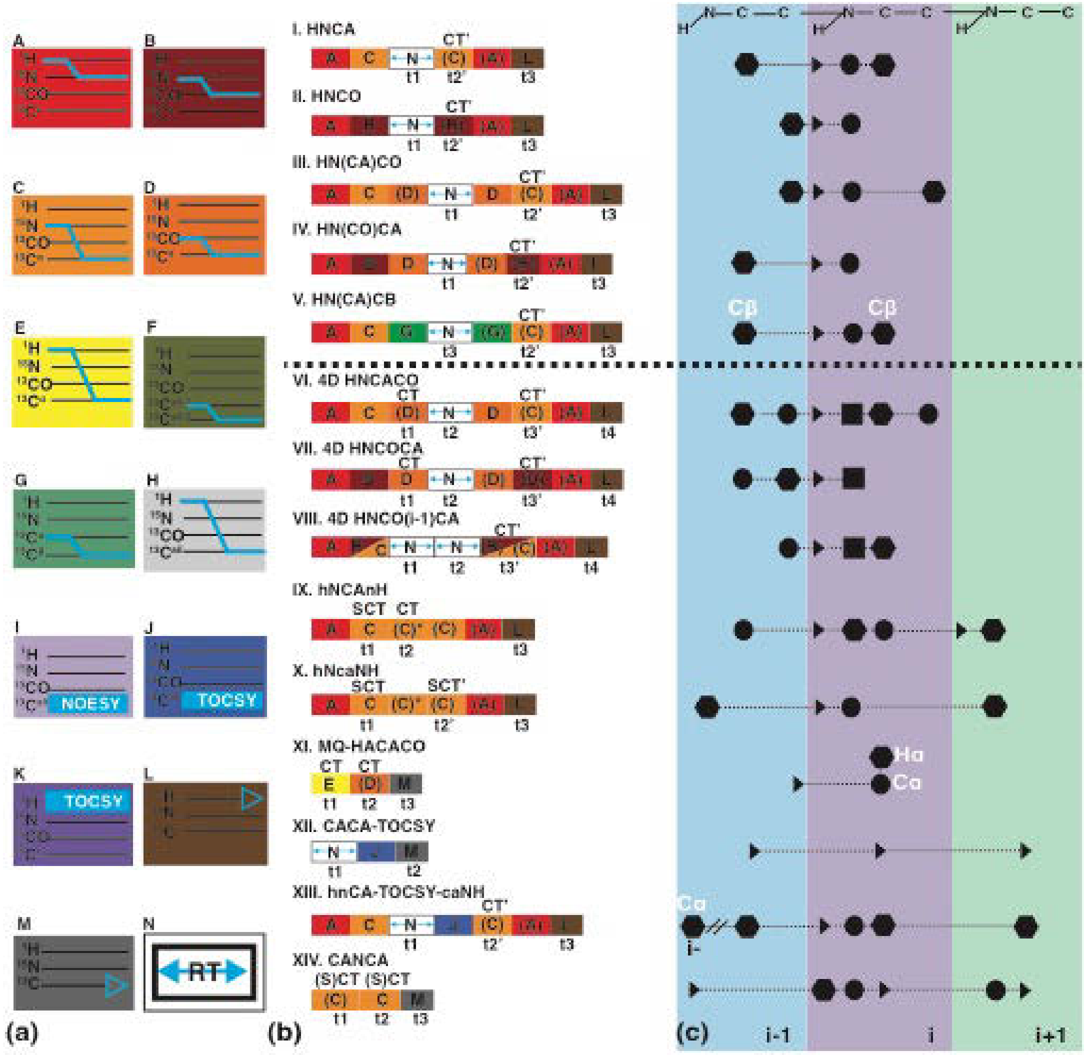


Figure 1.3 Graphical representation of pulse sequences used for backbone resonance assignment. Backbone pulse sequences (b) have been decomposed into a collection of basic building blocks (a) and paired with the information content used in resonance assignment (c). Panel (a) defines pulse sequence building blocks including coherence transfers (A-K), ^{1}H (L) or ^{13}C-detection (M), and chemical shift encoding in real time evolution (N). For example, in **A** represents an H→N INEPT, with $H_y \rightarrow 2H_zN_x$. (b) Backbone sequences are discussed in the text using the building blocks

defined in (a). Reverse transfers are denoted by parentheses; for example, ^{15}N transfer to ^{1}H is represented as (A) since it is the reverse of block A that represents transfer from ^{1}H to ^{15}N. Insertion of block N (real time evolution under chemical shift) implies evolution of the previously established coherence. (Semi) constant time is represented as (S)CT; CT indicates that the coherence established at the beginning of the block evolve in a CT manner. For example, the first CT period in VI (4D-HNCACO) corresponds to evolution under CO chemical shift. CT' represents evolution under chemical shift of the second coherence in the block; for instance in I (HNCA), the coherence transfer is from Cα to N, but the nitrogen chemical shift is encoded. An asterisk in hNCAnH and hNcaNH indicates a mixing period where C^{α} becomes antiphase with respect to both N_i and N_{i+1}. (c) Symbols indicate nuclei correlated in the corresponding experiment from panel (b). Hexagons, circles, and squares represent encoding during indirect dimensions (t_1, t_2, and t_3). Triangles denote detection.

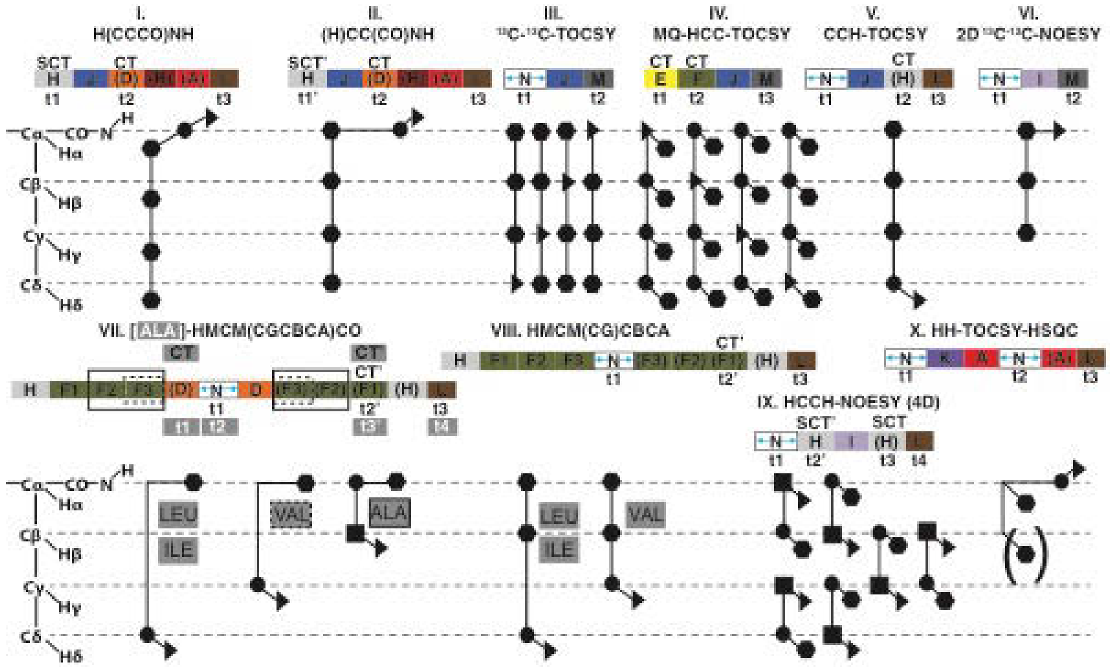


Figure 1.4 Graphical representation of pulse sequences used for side-chain resonance assignment. Building blocks are defined in Figure 1: panel (a). The nuclei correlated in each experiment are shown beneath each pulse sequence. Hexagons, circles, squares, and triangles represent encoding during t_1, t_2, t_3, and the detected dimensions, respectively. HMCM(CGCBCA)CO when shortened by two COSY periods becomes Val-HMCM(CBCA)CO, which lacks the F3 blocks (dotted box). The ALA-HMCBCACO is HMCM(CGCBCA)CO that is shortened by four COSY periods; it lacks the F2 and F3 blocks (solid lines). ALA-HMCBCACO has one more dimension than HMCBCACO, so its encoded dimensions are shown in white text. CT denotes constant time periods during which chemical shift is encoded on the coherence established at the beginning of the block, whereas, when the chemical shift encoding occurs on the second coherence of the block, CT' is used.

Chapter 2: The Yersiniabactin Synthetase Epimerization-Insert Domain: Modeling & Purification

Microbial supramolecular assemblies known as Nonribosomal Peptide Synthetases (NRPSs) produce a myriad of natural products including antibiotics, immunosuppressants and virulence factors[99]. Modular multi-domain architecture in NRPSs allows for so-called "assembly line" biosynthesis of these natural products. Each NRPS module incorporates and sometimes modifies one substrate[100–102]. Like many NRPSs, yersiniabactin synthetase (YS) creates heterocycles in its product, yersiniabactin, a crucial siderophore for *Yersinia pestis* growth[103–105]. YS employs domains that reflect a departure from classical NRPS assembly-line machinery including an unusual epimerization domain (E_A)[106]. Embedded in the primary sequence of an adenylation domain (A), which selects one substrate to be joined to the growing product chain, E_A controls the stereochemistry of an essential chiral center of yersiniabactin. This chapter contains a short introduction to nonribosomal peptides (NRPs) and to their natural synthesis, with a focus on the YS NRPS system. The chapter also describes how E_A domain boundaries were chosen and how E_A was purified. Finally, it demonstrates that E_A is structurally independent form the C-terminus of the A domain, in which it is embedded.

1. Background on non-ribosomal peptides and heterocycles

Microbes produce a myriad of secondary metabolites that are structurally and biologically diverse. Despite such remarkable structural and biological diversity, the so-called non-ribosomal peptides (NRPs) all contain peptide bonds that are enzymatically synthesized by non-ribosomal peptide synthetases (NRPSs), which are large (MDa) supramolecular assemblies that catalyze step-wise formation of NRPs. In *Figure 2.1*, six of these NRPs are shown. Of these six, epothilone and bleomycin are anti-cancer drugs, yersiniabactin and enterobactin are virulence factors (specifically, iron-chelating siderophores), and bacitracin and tyrocidine are antibiotics. Many promising therapeutic NRPs contain heterocyclic rings, which arise from tandem condensation and heterocylization of serines (or threonines) and cysteines to form oxazolines (methyl oxazolines) and thiazolines, respectively. Such heterocycles are intriguing targets for NRP research because of their stability and the ease of chemical diversification (such as methylation, oxidation, reduction, and epimerization). The subject of this thesis concerns the epimerization of the C9 center of yersiniabactin—specifically of its precursor, the intermediate hydroxyphenyl thiazoline (HPT) (*Figure 2.3*).

2. Introduction to non-ribosomal peptide synthetases

NRPS domains are grouped into so-called “modules,” which are sets of domains that are collectively responsible for the incorporation of a substrate (small molecule primary metabolite) into the growing NRP chain. (1) A freestanding phosphopantetheinyl transferase (PPTase) covalently modifies a serine of a small carrier protein (also called thiolation domain, T) with the phosphopantheine arm of Coenzyme A, converting the apo-T domain to its holo form. (2) Adenylation domains (A) select and activate substrates by creating adenylate intermediates. (3) The adenylated substrate intermediate reacts with the terminal thiol of the phosphopantheine arm, creating a metastable thioester linkage between the substrate and T domain, now called the loaded T-domain. (4) Condensation (C) domains catalyze peptide bond formation between thioester-linked substrates of adjacent modules, allowing the NRP chain to grow. Occasionally, cyclization (Cy) domains are used, which (as mentioned above) catalyze tandem condensation and heterocyclization of substrates. (5) Products are released from the NRPS machinery either by a thioesterase domain (TE) or by macrocyclization. In addition to the canonical domains described above, so-called “tailoring domains” within NRPSs allow for further chemical modifications of NRPs (such as methylation, oxidation, reduction, and epimerization). (*Figure 2.2*)

3. Introduction to yersiniabactin synthetase, yersiniabactin, and the epimerization domain of yersiniabactin synthetase

Yersiniabactin Synthetase, YS, produces the NRP yersiniabactin, which is a siderophore (iron-chelator) and is a crucial virulence factor of the etiological agent of plague, *Yersinia pestis*[107]. YS employs domains that reflect a departure from classical NRPS assembly-line machinery. Examples of these features include cyclization (Cy) domains that catalyze formation of aforementioned heterocycles, the E_A insert domain and an A domain that services two different T domains. The YS holoenzyme is made of four separate polypeptides[103] (*Figure 2.3*): an *in trans* A domain (YbtE), two high molecular weight proteins (HMWP1 and HMWP2), and an *in trans* reductase. *Figure 2.4* depicts the YS domains that are relevant to this thesis: **a**) YbtD[108], a PPtase modifies all T domains from apo to holo forms. **b**) YbtE, an in *trans* A domain, selects the substrate salicylate (sal) and transfers it to holo-T1[109]. Simultaneously, The YS A domain selects the substrate cysteine (cys) and loads holo-T2 (and holo-T3, not depicted). **c**) The Cy1 domain catalyzes (i) the condensation of salicylate and cysteine to form Sal-Cys-Ppant-T2 and (ii) the heterocyliczation of Sal-Cys-Ppant-T2 to form hydroxyphenyl-thiazoline-Ppant-T2 (HPT-Ppant-T2). YS A domain accepts only L-cysteine[106], which gives rise to the (*S*) stereoisomer of HPT at the C9 chiral center; however, the next Cy domain in the sequence (Cy2) accepts only (R)-HPT[106]. **d**) The combined activities of Cy2, T3, the hybrid NRPS/PKS HMWP1 and the reductase, YbtU complete the production of ybt.

Patel et al.[106] have investigated Pyochelin Synthetase (PS**)**, a *Pseudomonas aeruginosa* NRPS that produces pyochelin (pch, similar to ybt and also an HPT derivative) and has a similar arrangement of domains to YS. Both pch and the related enantio-pch (found in *Pseudomonas fluorescens*) are each stereospecific for both their intracellular transcription factors and their transmembrane receptors[110–112]. Patel et al. demonstrated that the 35 kDa E_A domain (*Figures 2.3* & *2.4*) was responsible for altering the stereochemistry of pch. PS E_A was determined to racemize linear (*S*)-Sal-Cys to a nearly equimolar racemic mixture of Sal-Cys, and Cy1 was shown to catalyze cyclization of both products to yield (*S*)-HPT and (*R*)-HPT. HPT ring opening does not spontaneously occur[104,113], so either of the two following strategies must be employed to remove unproductive (*S*)-HPT-Ppant-T2 and to prevent NRPS stalling: (**A**) E_A is able to racemize cyclized (*S*)-HPT, recognizing the heterocycles and/or (**B**) Cy1 in a reverse reaction, catalyzes linearization of the cyclized (*S*)-HPT to make (*S*)-Sal-Cys, recognized by E_A. These possible schemes are shown in *Figure 2.5*. Experiments described in this thesis are a component of a larger, overall goal of understanding, at an atomic level, the chemical substrates of E_A and the interactions of E_A with other YS domains (T1, Cy1, A), the three-dimensional structure of E_A and the molecular interactions that govern E_A embedding within the A domain.

4. Modeling of YS E_A based on alignment and topology of $A_{C\text{-term}}$ & demonstration of E_A's structural independence from $A_{C\text{-term}}$

A domains themselves are subdivided into N-terminal and C-terminal subdomains (*Figure 2.6*). E_A is inserted midway into the $A_{C\text{-term}}$ of the YS A domain. Although there are structures of 50 kDa epimerization domains that resemble condensation domains in sequence and protein fold, such epimerization domains are similar to E_A-

domains only in function, not in structure (based on predicted secondary structures, at this point) nor sequence. Since there are no structural representatives of E_A-domains, a crucial first step in elucidation of macromolecular structure is determination of domain boundaries. The Walsh lab originally estimated E_A domain boundaries as 1011-1348 based on rough sequence alignment to other A domains[106]. If we can map YS A properly, we can determine the insertion point of E_A. Eight A-domain structures were aligned and analyzed for structural similarity (Table 2.1). Using secondary structure topology maps, the relative arrangements of helices and sheets of the eight A domain structures were compared and mapped onto the PSI-PRED[114] secondary structure prediction output for the HMWP2 A-domain. We determined rough domain boundaries as 1008-1348 for E_A. This excised E_A was expressed as a fusion protein with thioredoxin in *E. coli* and purified to homogeneity. Likewise, the C-terminus of the A domain, $A_{C\text{-term}}$, was expressed and purified. HN-TROSY-HSQC spectra of both excised E_A and of $A_{C\text{-term}}$ were compared for chemical shift perturbation (CSP) analysis (*Figure 2.7 panel a*). CSPs were sorted by magnitude then assigned arbitrary peak numbers. The median, one-, and two-standard deviations of the data are plotted. Since within two standard deviations, the magnitude of CSPs is only ~ 0.06 ppm (*Figure 2.7 panel b*), we concluded that, when excised, E_A does not undergo any substantial structural rearrangement, and that therefore, it is justified to be studied in isolation from $A_{C\text{-term}}$.

We generated a homology model of E_A using the Phyre2 web portal[115] (*Figure 2.8*). Based on cursory inspection of the E_A Phyre2 model, we wondered whether E_A itself was comprised of two subdomains. The E_A Phyre2 model was used as input data for the Protein Peeling 2 server[116], which calculates theoretical divisions of subdomains based on spatial considerations (fewest number of intra-protein contacts). The residues in pink in *Figure 2.8* represent suggested cut sites for separation of E_A into N- and C-terminal subdomains. Six constructs were designed and tested for expression and solubility. (*Table 2.2*) Three of the 3 N-terminal subdomains (1009-1102, 1009-1114, & 1009-1135) and 2 of the 3 C-terminal subdomains (1102-1342 & 1114-1342) expressed solubly, but neither could be purified to sufficient quantity for NMR experiments (*Figure 2.9*). Given the discouraging results of C-terminal subdomain purification, we abandoned the notion of using subdomains and proceeded with our entire 341 residue construct (1007-1348).

Although the HN-TROSY-HSQC of E_A (*Figure 2.7 panel a* & *Figure 2.10, inlet*) exhibited exquisite spectral dispersion (indicative of a well-folded protein), a subset of peaks was markedly more intense than the others, possibly indicating disordered loops and/or N/C-terminal disordered linkers. We attempted to further truncate the C-terminus of E_A until the sample was no longer soluble. At the same time, buffers were exhaustively screened with respect to pH (pH 6.6, 6.8, 7.0, and 7.2), ionic strength (50 mM, 150 mM, 300 mM NaCl), and reducing agents (1 mM, 5 mM, 10 mM DTT ; 0.5 mM, 1 mM TCEP). In the end, we determined the proper domain boundaries of E_A would be 1008-1342 of HMWP2 and that data were best recorded in 20 mM Hepes 0.2 M NaCl 10 mM DTT pH 6.8 (at 298 K). The HN-TROSY-HSQC of (entirely protonated) ^{15}N-labeled EA is shown in *Figure 2.10*.

5. Detailed expression & purification methods

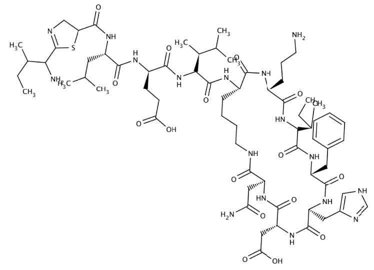
Basic expression and purification: The epimerization domain (E_A) of yersiniabactin synthetase corresponds to residues 1008-1342 of the *Yersinia pestis* protein HMWP2. $HMWP2_{1008-1342}$ was subcloned into the pET28a vector (Novagen, San Diego, CA), which has been modified to contain an additional N-terminal His_8 epitope, a thioredoxin (TRX) fusion protein, an internal His_6 site, and a TEV protease recognition: His_8-TRX-His_6-TEV-$HMWP2_{1008-1342}$. Heterologous protein expression was achieved in BL21(DE3) *E. coli* cells (Novagen) which were cultured in M9 medium: 6 g L^{-1} Na_2HPO4, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} NaCl, 2 mM $MgCl_2$, 0.1 mM $CaCl_2$, 7.5 µM thiamine, 1 g L^{-1} of NH_4Cl, 2 g L^{-1} glucose, pH 7.4. Bacteria were grown at 37˚C at 250 rpm until OD_{600} ~ 0.6 whereupon temperature was instantly lowered to 16˚C (ice bath). Protein expression was induced with 1 mM IPTG for 19 hours at 16˚C, 250 rpm. Cell pellets were harvested by centrifugation (5000 x *g,* 15 minutes, 4˚C), which were frozen at -80˚C until use. Cell pellets were resuspended in Lysis Buffer (50 mM Tris, 0.5 M NaCl, 30 mM Imidazole, 0.1% TritonX-100, pH 8.0 at 4˚C) and lysed using a Microfluidizer (Microfluidics Inc.). Lysates were clarified by centrifugation (27,000 *x* g, 30 minutes, 4˚C), filtered (0.22 µm) and passed onto a 5 mL HisTrap HP column (GE Healthcare) that had been equilibrated with Buffer A (50 mM Tris, 0.5 M NaCl, 30 mM Imidazole, pH 8.0 at 4˚C). Flow through (FT) and 80 mL of buffer A wash were collected separately, and TRX-E_A was eluted using a linear gradient of Buffer B (50 mM Tris, 0.5 M NaCl, 0.5 M Imidazole, pH 8.0 at 4˚C). TRX-E_A-containing fractions were pooled, 6 mL of 10 µM TEV protease (in

49.9% glycerol, 50% 50 mM Tris pH 7.5 (at room temperature) 0.2 M NaCl, 1 mM EDTA, 5 mM DTT, 0.1% TritonX-100) was added directly to the pooled eluate, and the pool (~35 mL) was dialyzed overnight at 4˚C against 2L of 50 mM Tris 0.2 M NaCl, 5 mM Imidazole, pH 8.0 at 4˚C. Dialysates were filtered (0.22 μm) and passed through a 5 mL HisTrap HP column that had been equilibrated with dialysis buffer; FT was collected into 10 mL dialysis buffer supplemented with 10 mM DTT, concentrated to 2 mL, and buffer exchanged twice (Using Amcon 10 kDa molecular weight cutoff filters, Millipore: 15 mL → 1.0 mL) into 20 mM Hepes, 0.2 M NaCl, 10 mM DTT, pH 6.8 at 4˚C. Concentrate was then separated using size exclusion chromatography with a Superdex 75 16/60 column (GE Healthcare). Fractions giving rise to a single, monodispersed peak were concentrated to 0.6 mM. Typical yields were 600 μL of 0.6 mM E_A per L of M9 culture. In NMR samples, specimens contained 5% ^2H_2O for lock (and $^{15}NH_4Cl$ was used to isotopically label the protein) and 0.2 mM DSS (prepared in ^2H_2O).

6. Conclusions & Future Directions

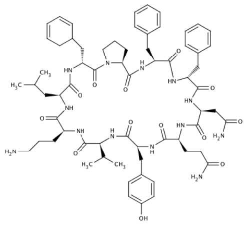
Section 4 described that rough domain boundaries for E_A were determined based on predicted secondary structure matching of its surrounding $A_{C\text{-term}}$. We determined conditions by which soluble well-folded E_A could be produced (as determined by size exclusion chromatography and $^1H^{15}N$-TROSY-HSQC). We then compared $^1H^{15}N$-TROSY-HSQC of E_A to $A_{C\text{-term}}$, in which it is embedded, which revealed that there were no major structural changes of E_A when it is excised from $A_{C\text{-term}}$ (*Figure 2.7*). We then attempted to determine possible N- and C-terminal subdomains of E_A (which we hypothesized based on observations from limited proteolysis, variations in peak intensity, and homology modeling [*Figure 2.8*]). Every attempt to divide E_A into subdomains resulted in unstable samples that could not be purified to sufficient quantity for a structural biology NMR project (*Figure 2.9)*. We performed buffer screens on E_A and settled upon the buffer (20 mM Hepes, 0.2 M NaCl, 10 mM DTT, pH 6.8 at room temperature) because it gave the most uniform $^1H^{15}N$-TROSY-HSQC and allowed the protein to remain soluble for approximately two weeks at room temperature and at 600 μM.

We chose to initiate a daunting NMR structural biology project. Structure determination by NMR spectroscopy was briefly mentioned in *Chapter 1, Section 3.ii.* A schematic of the roadmap of protein NMR structure determination is shown in *Figure 2.11*. First, protein must be expressed and purified in sufficient yield (> 300 μL of > 300 μM), it must be stable (free of aggregation and precipitation) for at least two weeks at room temperature, and its ${}^{1}H^{15}N$-TROSY-HSQC spectrum should appear well dispersed with signals of uniform intensity. Once buffer conditions are suitable, protein is grown with isotopic enrichment of ${}^{1}H$, ${}^{15}N$, and ${}^{13}C$, allowing for a suite of triple resonance experiments to be collected. Spectroscopic signals are paired with nuclei of the protein backbone in a process known as sequential resonance assignment. For large proteins, assignment will be thwarted because of loss of connectivity (in which case longer experiments should be performed to provide i-2, i-1, i, i+1, and i+2 correlations), loss of sensitivity (in which case more scans should be collected or higher protein concentration should be used), or because of slow back exchange of amide deuterons for protons (in which case equilibrium refolding in protonated solvent or ${}^{13}C$ detected experiments can be used). Once backbone assignment has been completed, side-chain assignment is performed. With C^{α} and C^{β} chemical shifts in hand, the TALOS[117], TALOS+[60], or TALOSN[118] software may be used to determine torsional angles (and ergo secondary structure). Finally, distance constraints are collected with 3D- (or 4D-)${}^{13}C$-edited, ${}^{15}N$-edited, or time-shared NOESY. Distance constraint information and torsional angle restraints (from C^{α} and C^{β} chemical shifts) are used as constraints during energy minimization of an ensemble of random, extended polypeptidic chains to determine a bundle of structures. Violations are inspected and distance restraints are re-inspected until the structural bundle is well converged (low RMSD) and satisfies an energy force field (low target function).



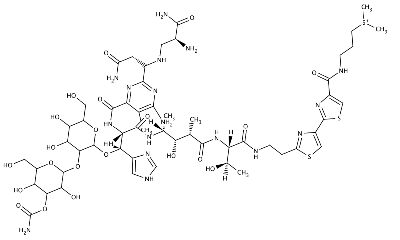
antibiotic

bacitracin



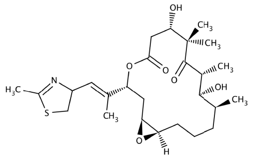
antibiotic

tyrocidine



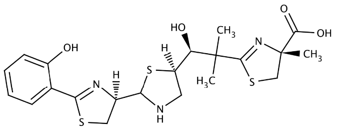
anti-cancer

bleomycin



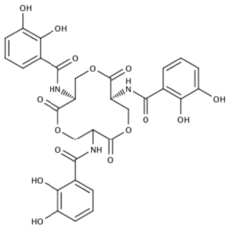
anti-cancer

epothilone



siderophore

yersiniabactin



siderophore

enterobactin

Figure 2.1 Medically relevant non-ribosomal peptides (NRPs). Six NRPs produced by nonribosomal peptide synthetases. Anti-cancer agents (epothilone and bleomycin) are shown in blue, virulence factors are shown in green (yersiniabactin and enterobactin, iron-chelating siderophores), and antibiotics (bacitracin and tyrocidine) are shown in purple. Some of the NRPs shown here contain heterocyclic rings.

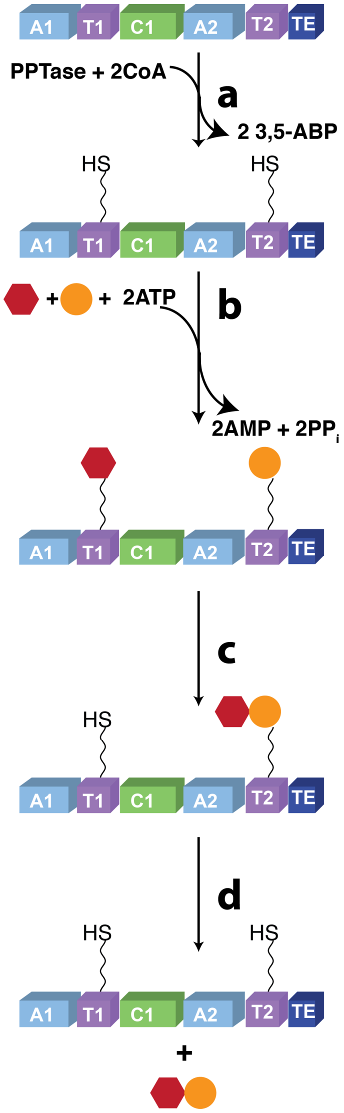
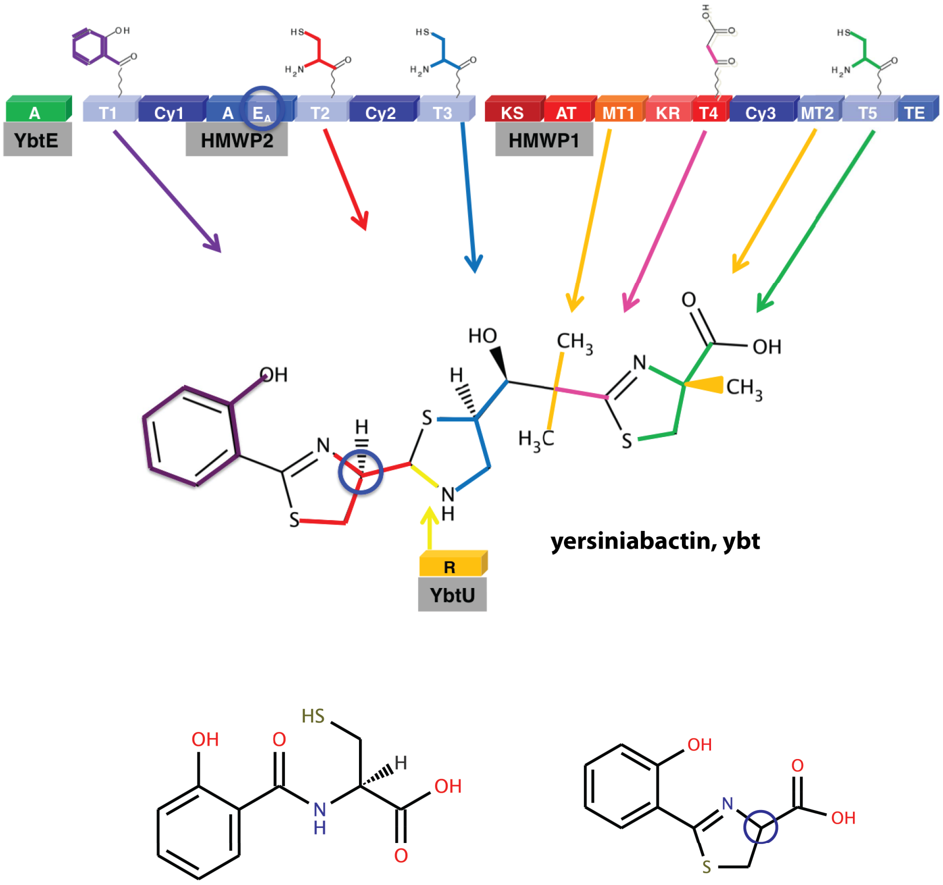


Figure 2.2 Assembly line synthesis by Nonribosomal Peptide Synthetases, NRPSs. A schematic representation of a general NRPS is shown in the figure. Catalytic domains are represented with cubes. Adenylation (A) domains are shown in blue, thiolation (T, also called carrier protein) domains are shown in purple, a condensation domain (C) is shown in green, and a thioesterase (TE) domain is shown in dark blue. Synthesis occurs as follows: (a) a phosphopantetheinyl transferase (PPTase) transfers the phosphopantethine arm (wavy line) derived from CoA to T1 and T2, with AMP as a byproduct. (b) The adenylation domains (A1 and A2) select

substrates (hexagon and circle, respectively), form adenylyl intermediates, then finally transfer the substrates to T1 and T2, forming thioester linkages. (c) The condensation domain (C1) catalyzes peptide bond formation between the two substrates. (d) Finally, a thioesterase catalyzes hydrolysis of the substrates from the NRPS machinery, liberating the substrates.



salicylate-cysteine, Sal-Cys

hydroxyphenyl thiazoline, HPT

Figure 2.3 Yersiniabactin Synthetase (YS), yersiniabactin (ybt), and hydroxyphenyl thiazoline (HPT). YS is made of four separate polypeptide chains: YbtE an *in trans* A domain, High Molecular Weight Proteins 2 & 1 (HMWP2 & HMWP1), and YbtU, an *in trans* reductase. YS produces the iron-chelating siderophore, yersiniabactin (ybt). The chiral center of ybt (C9) that is racemized by the epimerization domain, E_A, is highlighted with a blue circle. Salicylate-cysteine (Sal-Cys) and hydroxyphenyl thiazoline (HPT) are intermediates in the biosynthesis of ybt.

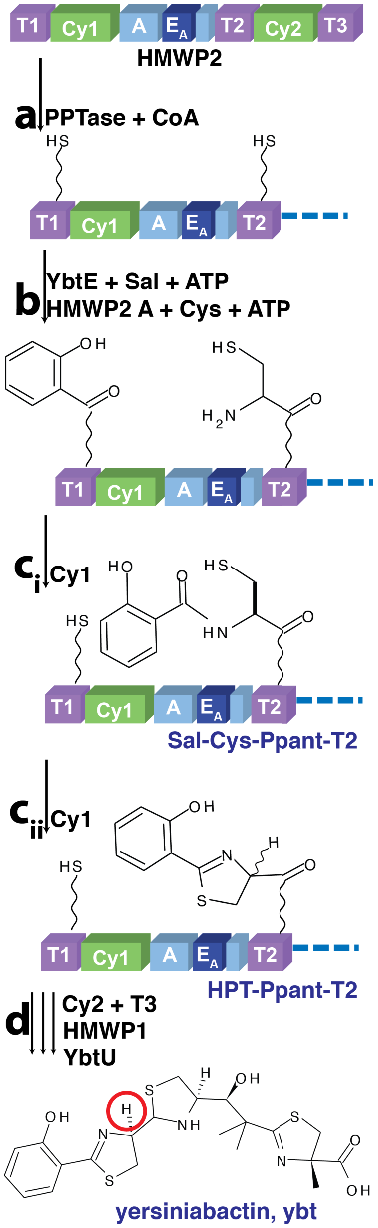


Figure 2.4 Yersiniabactin Synthetase domains relevant to this thesis and biosynthesis of HPT, ybt. (a) thiolation domains T1 and T2 of HMWP2 are phosphopantetheinylated (b) YbtE adenylates salicylate (Sal) and passes it to T1 and the A domain of HMWP2 adenylates cysteine (Cys) and passes it to T2. (c_i) Cy1 first catalyses peptide bond formation between Sal and Cys to form the Sal-Cys intermediate, then (c_{ii}) cyclizes Sal-Cys to HPT. (d) Further catalytic actions of other YS domains lead to the formation of ybt.

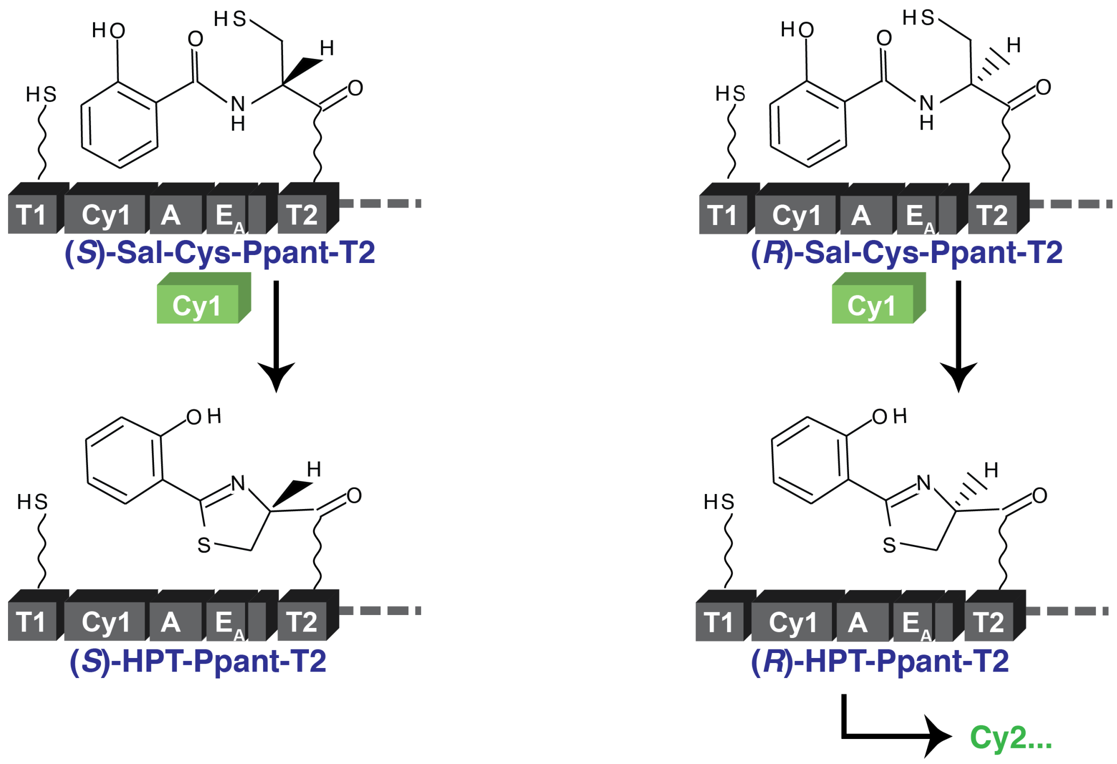


Figure 2.5 Possible mechanisms of E$_A$-catalyzed racemization of HPT. It is unclear whether YS E$_A$ (i) directly racemizes the linear Sal-Cys, which is then converted from (*R*)-Sal-Cys-Ppant-T2 into (R)-HPT-Ppant-T2, (ii) racemizes the cyclized (*S*)-HPT-Ppant-T2, or (iii) can reacemize both substrates. Importantly, the only substrate recognized by the downstream Cy2 domain is (*R*)-HPT-Ppant-T2. Therefore, interfering with the activity of E$_A$ may be a useful strategy to prevent production of the ybt virulence factor, indicating that E$_A$ could be an attractive, drugable target.

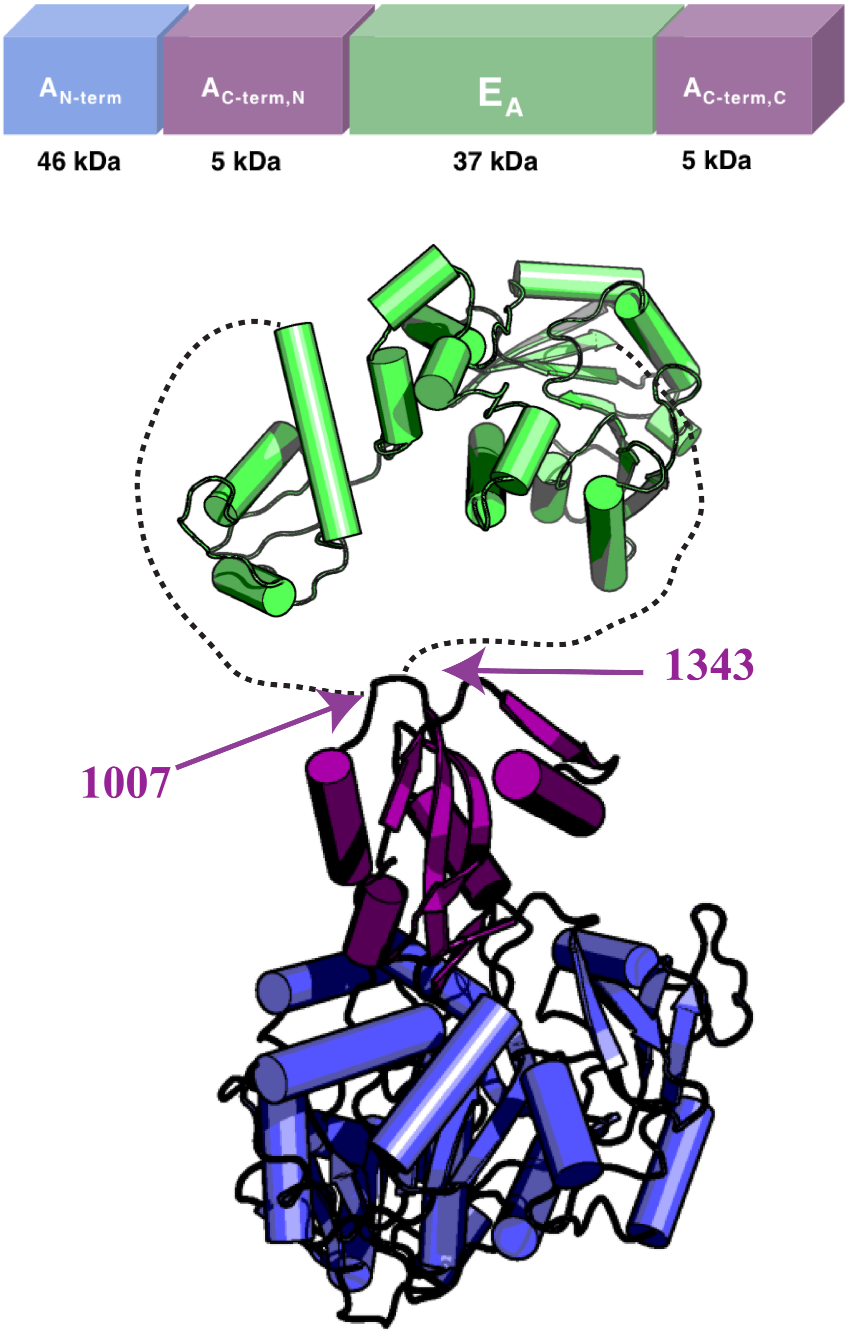


Figure 2.6 Anatomy of E_A & $A_{C\text{-term}}$. A domains are composed of N and C-terminal subdomains of ~ 45 kDa and ~ 10 kDa, respectively. E_A is inserted midway into the sequence of $A_{C\text{-term}}$. Since the structure of the YS A domain has not been determined, an example A domain structure from Enterobactin (EntE) is shown (PDB code

3RG2[119]), for illustration. This A-domain's **N-** and **C-**termini are shown in **blue** and **purple**, respectively. Although this example A-domain does not contain an insert E_A, the position on this A domain's C-terminus that corresponds to the insertion of E_A in YS A domain is represented with two purple arrows. E_A is represented by the homology model described in *Figure 2.8*.

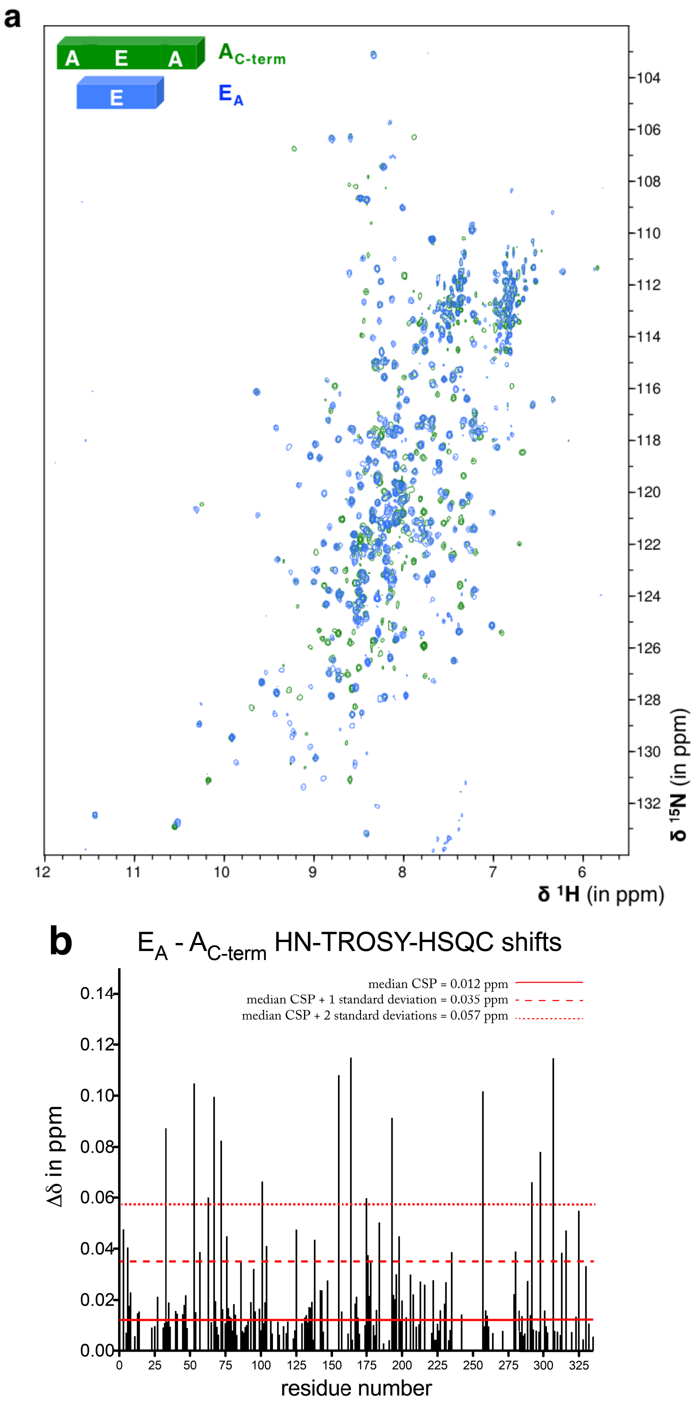


Figure 2.7 E_A is structurally independent from $A_{C\text{-term}}$. **(a)** Overlay of HN-TROSY-HSQCs of $A_{C\text{-term}}$ (green) with E_A (blue). Samples were prepared in 20 mM Sodium Phosphate, 0.15 M NaCl, 10 mM DTT, pH 6.6 and concentrated to ~300 µM,

5% of sample was composed of ^2H_2O for lock. 16 transients were recorded with a 1 second recycling delay. 512 and 128 complex points were recorded in the ^{1}H and ^{15}N dimensions, respectively. Data were collected at 600 MHz at 298 K. **(b)** Peak shifts were quantified as $\Delta\delta \text{ (in ppm)} = \sqrt{\left(\frac{\Delta\delta_N}{5}\right)^2 + (\Delta\delta_{HN})^2}$ and plotted as a function of arbitrary peak number. The median chemical shift perturbation is shown with a solid red line, the median + 1 standard deviation is shown in the coarsely dotted red line, and the median + 2 standard deviations is shown as the finely dotted red line.

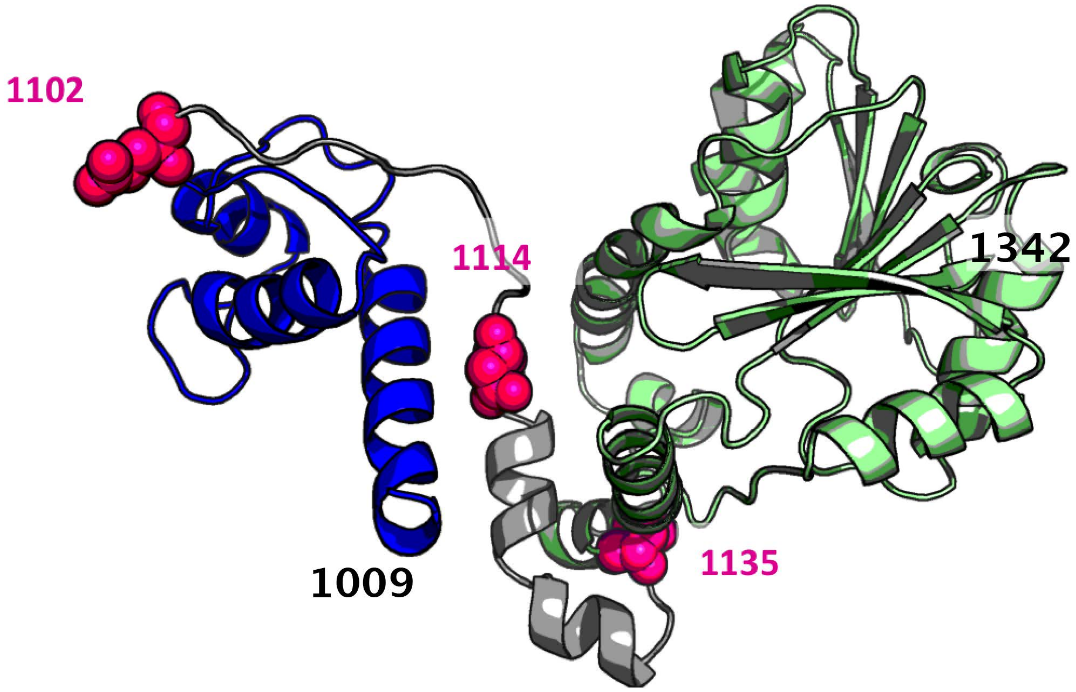


Figure 2.8 Phyre2[1] homology model of E_A. Two potential subdomains are shown in blue and green, as determined by the Protein Peeling2 web server[116]. Residues used as "cut points" to separate the potential $E_A^{\text{N-terminal}}$ subdomain from the $E_A^{\text{C-terminal}}$ are shown in pink (see *Table 2.2*).

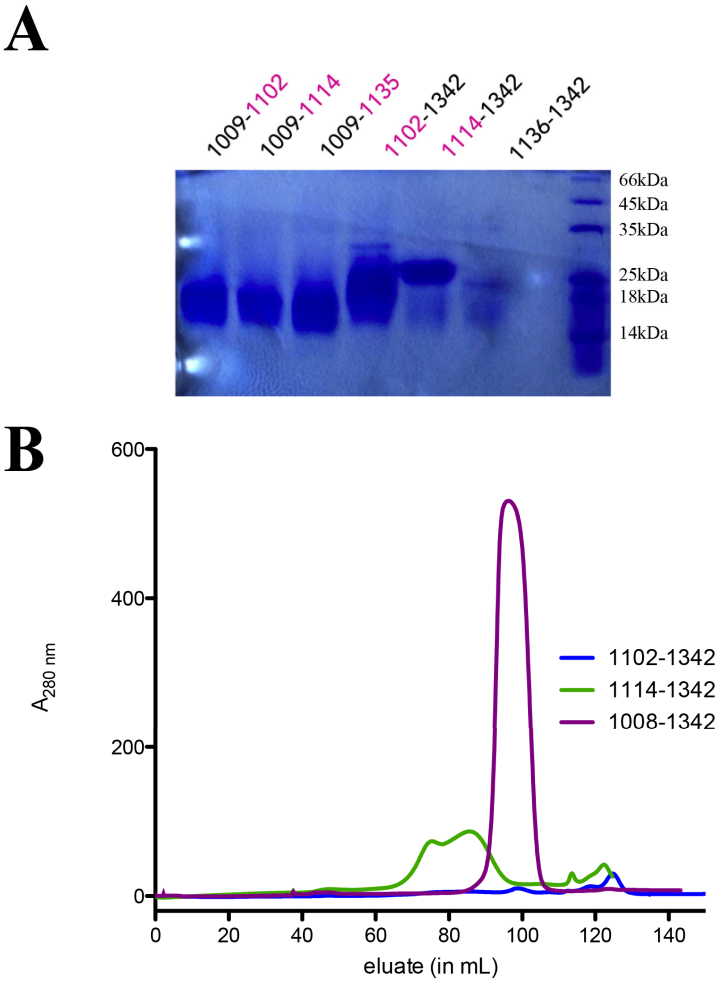


Figure 2.9 Expression & Purification of putative E$_A$ N- & C-terminal subdomains. (A) SDS-PAGE of soluble fraction of TRX-E$_A$-subdomain constructs reveals that 5 of the 6 putative subdomain constructs expressed well in *E. coli* using the TRX-E$_A$ expression protocol. (B) Constructs 1102-1342 (**blue**) and 1114-1342 (**green**) were purified following the same protocol as E$_A$. The final purification products were subjected to size exclusion chromatography (using Superdex 200 resin). These products did not elute at the anticipated volume (i.e. > 90 mL), and the protein recovery was insufficient (1008-1342 expressed with the same amount of bacterial culture is shown in **purple** for a comparison).

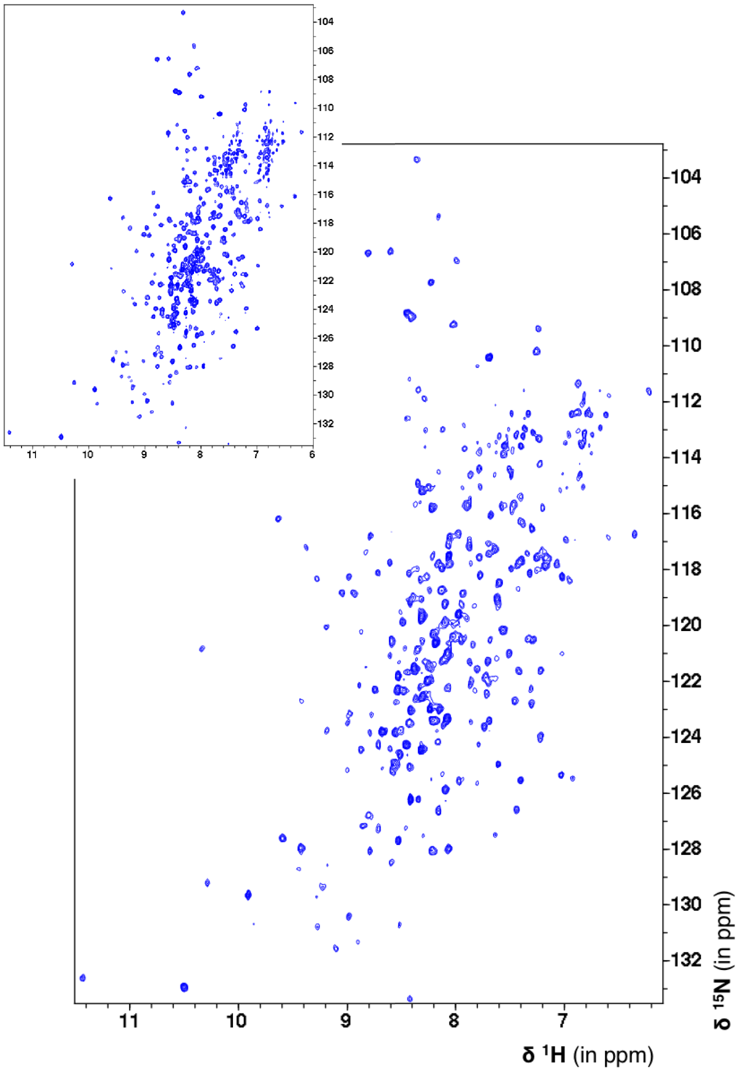


Figure 2.10 HN-TROSY-HSQC of E_A under optimized buffer conditions and domain boundaries. *(Larger spectrum)* Optimized construct and buffer conditions 600 µM E_A (residues 1008-1342 of HMWP2) was expressed and purified as described in *Section 5*. NMR buffer conditions: 20 mM Hepes 0.2 M NaCl 10 mM DTT pH 6.8 (at room temperature), 5% ^2H_2O. This spectrum exhibits excellent spectral dispersion, indicating that the sample is well folded and suitable for further NMR studies. (*Inlet*) Starting construct and buffer conditions 600 µM E_A (residues 1007-1348) expressed and purified as described in *Section 5*. NMR buffer conditions: 20 mM Sodium Phosphate 0.2 NaCl 5 mM DTT pH 6.8 (at room

temperature) 5% ^2H_2O. In both spectra, 16 transients were recorded (using a 1 second recycling delay) on a 600 MHz spectrometer at 298 K. 512 complex points and 128 complex points were recorded in the ^1H and ^{15}N dimensions, respectively. Altering the construct boundaries and the buffer conditions substantially improved the quality of NMR spectra.

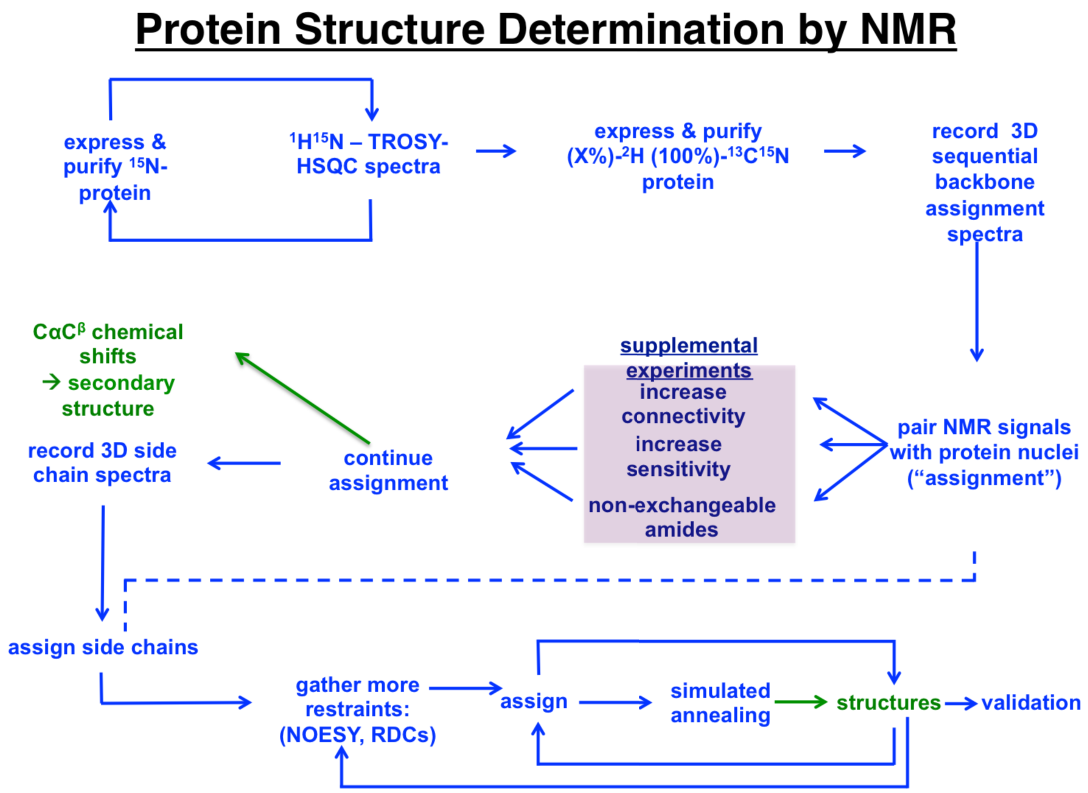


Figure 2.11 Protein Structure Determination by NMR Roadmap.

Determining a protein structure with NMR spectroscopy is a multi-step process, described in text.

Table 2.1 Structures of A-domains used in assessing E_A domain boundaries

PDB accession code	Literature citation
3GPC	Kochan G. et al. *Journal of Molecular Biology 2009*[120]
1AMU	Conti E. et al. *EMBO Journal 1997*[121]
3CW9	Reger A. et al. *Biochemistry* 2008[122]
3E7X	Yonus H. et al. *Journal of Biologial Chemistry* 2008[123]
3IPL	Patskovsky Y. et al. (unpublished, Protein Structure Initiative)
3A9V	Hu Y. et al. *Plant Cell* 2010[124]
3IVR	Patskovsky Y. et al. (unpublished, Protein Structure Initiative)
3ITE	Verne Lee T. et al. *Journal of Biological Chemistry* 2010[125]

Table 2.2 Structures of A-domains used in assessing E_A domain boundaries

Construct	Experimental observations
1009-1102	Expressed as TRX-fusion protein, not soluble when purified
1009-1114	Expressed as TRX-fusion protein, not soluble when purified
1009-1135	Expressed as TRX-fusion protein, not soluble when purified
1102-1342	Expressed as TRX-fusion protein, not soluble when purified
1114-1342	Expressed as TRX-fusion protein, not soluble when purified
1136-1342	No expression

Chapter 3: Assignment of E_A using partial deuteration, non-uniform sampling, equilibrium protein refolding, covariance NMR, and ILV-labeling

A modified portion of this text has been published in the *Journal of the American Chemical Society*.

1. Backbone Assignment using 70-DCN sample

As mentioned in the introduction (*Chapter 1, Section 4.vi*), partially deuterated protein samples can be tremendously useful in the pursuit of NMR studies of larger biological macromolecules. The choice of amount of protein deuteration (that is, the percentage of growth solvent that is ^2H_2O relative to ^1H_2O) is typically empirically determined. Increasing the percentage of ^2H_2O lessens the amount of dipole-dipole mediated relaxation (since ^2H has ~ 1/6 the gyromagnetic ratio of ^1H). On the other hand, increasing the ratio of ^2H_2O relative to ^1H_2O reduces the number of nOe cross-peaks that are crucial for structure calculation. For backbone resonance assignment of E_A, we prepare E_A by expressing E_A as described in the *Chapter 2*, *Section 5* with the following exceptions: growth medium solvent was composed of 70% ^2H_2O and 30% ^1H_2O ($pH^{meter} = 7.0$), and ^{15}N and ^{13}C were exogenously supplied with $^{15}NH_4Cl$ and $^1H^{13}C$-glucose.

Because protein expression occurred in cells grown in 30% ^1H_2O, the 37 kDa E_A domain contained enough protons to be subject to sensitivity losses due to effective relaxation processes; whereas, complete deuteration (perdeuteration) would show superior sensitivity (*Chapter 3, Section 6*). To further combat relaxation, we employed the TROSY technique on the following triple resonance experiments: TROSY-HNCO, TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HN(CA)CO, TROSY-HN(CA)CB, and NOESY-TROSY-HSQC. These sequences were modified (as described in *Chapter 1, Section 4.vi*) to refocus coherence evolution under both ^1H-^{13}C and ^2H-^{13}C scalar couplings. In all of our sequences, ST2-PT TROSY elements are used in concert with N→CA/CO concatenation (shortening the sequence by 5.5. ms)[92]. Sensitivity of these experiments was still relatively poor, so we used NUS to bolster sensitivity (with exception of the NOESY-TROSY-HSQC). By sampling 15% of indirect dimension points (while still keeping N_{max} within conventional resolution parameters), we were able to increase the signal-to-noise of our experiments within a conventional amount of experimental time. Data acquisition parameters are shown in *Table 3.1*. Although we chose to employ NUS to facilitate averaging of more signals and rescue sensitivity rather than access ultra-high resolution, we nevertheless targeted resolutions beyond the limitations of the ^{15}N constant time period. To bypass this limitation, our TROSY sequences feature semi-constant time in the ^{15}N dimension[126]. Similarly, our TROSY-H(CCCO)NH and TROSY-(H)C(CCO)NH feature semi-constant time periods in the $^1H_{indirect}$ and ^{13}C dimensions, respectively (*Figure 3.1*). Data were reconstructed using istHMS as described in the Introduction (*Chapter 1, Section 3.c*) and representative 2D projections of HNCO (most sensitive experiment), HN(CA)CB (least sensitive experiment), and HN(CA)CO (moderately sensitive experiment) are shown in *Figure 3.2* to demonstrate the impressive sensitivity and quality of 3D data collection.

Once data collection and reconstruction was achieved, we moved directly to backbone assignment. Assignment of NMR resonances relies on identifying (H,N) correlations that belong to sequential residues in the polypeptidic chain. Two distinct types of 3D spectra convey this information. In the first type, an additional dimension encodes carbon chemical shifts of both the same and the preceding residue (referred to as Intra-3D to emphasize the former correlation). The second type reports only carbon chemical shift of preceding residues (Seq-3D). The assignment procedure consists of identifying (H(i+1),N(i+1),C(i)) correlations for residue i in the Seq-3D that feature carbon shifts matching that of a correlation (H(i), N(i), C(i)) found in the Intra-3D. This process is performed using C^{α} (with HNCA for Intra-3D and HN(CO)CA for Seq-3D), CO (HN(CA)CO and HNCO), and when possible, C^{β} (HN(CA)CB and HN(COCA)CB) chemical shifts. The practical implementation of the procedure comprises a series of steps. First, (H,N,C) correlations are identified by peak picking. Next, H/C (or N/C) strips are generated for each peak in each spectrum. The strip of a target residue is selected in Intra-3D and a software package sorts all strips of Seq-3D according to the difference in carbon frequencies as determined by peak picking (strip matching). In general, the procedure requires simultaneous analysis of different carbons (C^{α}, CO, and C^{β}) to help identify true sequential residues and eliminate accidental degeneracies in carbon frequencies. Using NUS-collected

3D data, approximately 1/3 of backbone residues could be assigned using the techniques described as shown in *Figure 3.3*, where yellow represents assigned backbone residues, white represents unassigned backbone residues, and red represents prolines that lack H^N.

2. Covariance NMR spectroscopy

Clearly, the assignment procedure relies on the accuracy of peak picking, which greatly deteriorates in the presence of frequency degeneracies. Thus, unpicked correlations will not be represented during strip matching. In addition, carbon frequencies of different spectra can be mispaired with (H,N) correlations that overlap; for example the C^{α} of residue *i* could be paired with the CO of residue *j*. In such a case, strip matching will either be unsuccessful or worse, erroneous. To overcome the limitations of pick peaking, we have designed spectral manipulations that, in essence, replace this convoluted assignment procedure with a simple inspection of four 3D correlation maps. Moreover, each map reports on the combined sequential information contained within *all* pairs of Intra-3D and Seq-3D spectra. The four correlation maps provide correlations of the form (H(*i*), N(*i*), H(*i*+1)), (H(*i*), N(*i*), N(*i*+1)) (H(*i*), N(*i*), H(*i*-1)), and (H(*i*), N(*i*), N(*i*-1)) and thus permit direct identification of sequential residues in (H,N) correlation maps. The method employs covariance[127–130], albeit with spectra suitably modified to minimize artifacts. Covariance and related methods were suggested as tools to help protein assignment by creating novel correlations[131–133], but artifacts have limited applications to small proteins where such artifacts can be identified. Another elegant solution was tailored to sequential assignment[134], but it required peak picking and is hence vulnerable to its associated limitations. Overall, covariance methods have not been widely adopted for resonance assignment. Here, sequential correlation maps with minimal artifacts are obtained by (i) taking a spectral derivative prior to covariance between pairs of Intra- and Seq-spectra and (ii) multiplying the resulting covariance correlation maps to combine the information provided separately by different carbon dimensions into a single spectrum.

Covariance NMR can be used to provide a spectral representation of the sequential assignment procedure; however preliminary treatment of the original spectra is needed to minimize artifacts. To identify and overcome shortfalls of covariance NMR in the presence of near degenerate frequencies, we will first reformulate the sequential assignment procedure in a context that over-represents overlap: "Amongst all (H,C$^{\alpha}$) correlations in HN(CO)CA, find the one that possesses a C$^{\alpha}$ frequency matching the observed C$^{\alpha}$ in HNCA for an (H,N) correlation", and likewise for all pairs of spectra. The mathematical formulation of this procedure consists of calculating the covariance matrix between the H/C projection of HN(CO)CA, referred to as 2D-H(NCO)CA, and each H/C plane of HNCA (for all nitrogen indices). Using the formalism of Brüschweiler and co-workers[129,135,136], the following 3D array can be constructed:

$$HNH_sca(a,b,c) = \frac{1}{D-1}\sum_{d=1}^{D} \widetilde{HNCA}(a,b,d) \cdot H(\widetilde{NCO})CA(c,d), \qquad (1)$$

The symbol "~" indicates that the means along the carbon dimensions have been subtracted from the spectra[2]. Indices *a* and *c* represent the HNCA and 2D-H(NCO)CA

^{1}H dimensions, respectively; b is the index along the HNCA ^{15}N dimension, and d is the common index along the ^{13}C dimensions of both spectra (each with D points). The resultant 3D spectrum, HNH$_s$ca, correlates (H,N) correlations of HNCA with sequential H$_s$ resonances of 2D-H(NCO)CA. HNH$_s$ca provides correlations (H(i), N(i) , H(i+1)) and is an array of covariance matrices HH$_s$ dispersed along a nitrogen dimension. Unfortunately, in practice, a multitude of false correlations appear in such a correlation map. To identify the origin of these artifacts and to design a solution, we reformulate the mathematics of covariance NMR into two district steps: the element-wise product of two C$^\alpha$ vectors and subsequent summation over the elements of the resulting vector. First we define a vector

$$\overrightarrow{v_{a,c}}(d) = \overrightarrow{ca}_a(d) \odot \overrightarrow{ca_{-1}}_c(d) \tag*{(2),}$$

where $\overrightarrow{ca}_a$ and $\overrightarrow{ca_{-1}}_c$ are vectors representing 1D C$^\alpha$ traces at ^{1}H frequencies defined by the index a in HNCA and c in 2D-H(NCO)CA. Here $\odot$ denotes the element-wise product and the symbols "~" have been omitted for clarity. Each point (a,c) in the plane HH$_s$ is proportional to the sum of the elements of the vector $\overrightarrow{v_{a,c}}$:

$$HH_s(a,c) = \frac{1}{D-1} \sum_{d=1}^{D} \overrightarrow{v_{a,c}}(d) \tag{3}$$

By observing the individual C$^\alpha$ vectors and their associated element-wise products $\vec{v}$ prior to summation, we can discern the origin of artifacts in HH$_s$ that have plagued covariance NMR thus far.

Figure 3.4 uses simulated data to demonstrate the source of artifacts in covariance NMR spectra. *Figure 3.4 panels a & b* display the same vector $\overrightarrow{ca}_a$ at an index H(i) =

a in ^{1}H of HNCA. *Figure 3.4 panel c* displays a vector $\overrightarrow{ca}_a{}^{*}{}_{-1_c^*}$ that contains the true sequential peak at index H(i+1) = c^* in 2D-H(NCO)CA, while *Figure 3.4 panel d* displays $\overrightarrow{ca}_a{}^{X}{}_{-1_c^X}$ containing a nearly degenerate C$^\alpha$ peak at index H(i+1) = c^X. The element-wise products of $\overrightarrow{ca}_a$ with $\overrightarrow{ca}_a{}^{*}{}_{-1_c^*}$ ($\vec{v}^*$) and $\overrightarrow{ca}_a$ with $\overrightarrow{ca}_a{}^{X}{}_{-1_c^X}$ ($\vec{v}^X$) are shown in *Figure 3.5 panels e and f*, respectively. Summing the vectors $\vec{v}^*$and $\vec{v}^X$ provides the amplitudes of HH$_s$ at indices (a,c^*) and (a,c^X) in *Figure 3.4 panel m*. We can see a false correlation resulting from partial overlap in the C$^\alpha$ dimension. This artifact can be reduced by taking the derivative along the C$^\alpha$ dimensions prior to covariance (*Figure 3.4 panels g—j*). In this case, $\vec{v}^{(\prime)*}$ now contains only positive elements (*Figure 3.4 panel k*), while $\vec{v}^{(\prime)X}$ contains both positive and negative elements due to the mismatched inflection points in $\overrightarrow{ca}_a{}'$ and $\overrightarrow{ca}_a{}'^{X}{}_{-1^X}$ (*Figure 3.4 panel l*). Summing $\vec{v}^{(\prime)*}$ results in positive correlation at index (a,c^*) in *Figure 3.4 panel n*, whereas the sum of $\vec{v}^{(\prime)X}$ gives zero amplitude at index (a,c^X). Here, the degree of C$^\alpha$ frequency degeneracy was chosen to completely suppress artifacts when using spectral derivatives. Strong degeneracy would result in positive yet reduced artifacts, while weaker degeneracy would create negative artifacts that can safely be ignored.

Figures 3.5 and *3.6* illustrate experimentally the effectiveness of artifact suppression in covariance matrices when using derivatives of original spectra. *Figure 3.5* shows $\overrightarrow{ca}$ and $\overrightarrow{ca}_{-1}$vectors as well as their element-wise products $\vec{v}$, and *Figure 3.6 panels a* and *b* show traces from the covariance matrix HH$_s$. Although the vector $\overrightarrow{ca}$ shown in

Figure 3.5 panel a should only correlate with $\overrightarrow{ca}_{-1}^{*}$ (*Figure 3.5 panel b*), it also correlates, among others, with $\overrightarrow{ca}_{-1}^{X}$ (*Figure 3.5 panel c*). Both vectors $\vec{v}^{*}$and $\vec{v}^{X}$ (*Figure 3.5 panels d* and *e*) have only positive elements that, after summation, give rise to the signals labeled * and X in *Figure 3.6*. Results are improved if the derivatives of the vectors $\overrightarrow{ca}$ and $\overrightarrow{ca}_{-1}$are used for covariance analysis (*Figure 3.5 f—h*). After element-wise multiplication (*Figure 3.5 panels i* and *j*) and summation, the amplitude of the artifact is either reduced or becomes negative in the covariance matrix (*Figure 3.6 panel b*, signal labeled *X*). Thus, true sequential correlations can be distinguished to a large extent from contributions of residues with carbons of nearly identical frequencies.

A single COSCOM (covariance sequential correlation maps) conveys information obtained with four separate original spectra. The traditional sequential assignment procedure requires that C^{α} and CO strips, for example, be analyzed in parallel to delineate accidental frequency degeneracies from true sequential correlations. The COSCOM procedure applied to C^{α} in the previous paragraph can also be applied to HN(CA)CO and HNCO to produce $HNH_{s}co$ spectra (*Figure 3.6 panels c* and *d*). Because $HNH_{s}ca$ and $HNH_{s}co$ both provide sequential correlations along a common proton dimension, placing the same spectral region side-by-side (or overlaying them) readily identifies common sequential correlations. Alternatively, the sequential information contained in each COSCOM spectrum can be combined via element-wise multiplication into a single spectrum, thus permitting further reduction in artifacts due to the destructive interference of erroneous correlations. Indeed, *Figure 3.6 panels e* and *f* show that multiplication of HNH$_s$ca and HNH$_s$co to produce HNH$_s$caco removes a majority of the erroneous correlations that resulted from accidental (near) degeneracies in C$^\alpha$ and CO carbon frequencies. Without using spectral derivatives, only three sequential proton candidates remain in HNH$_s$caco (*Figure 3.6 panel* e). However, when taking the derivative prior to covariance only a single correlation remains. The other two correlations are severely damped, reflecting that they originate from partial overlap in ^{13}C signals, and the true sequential correlation is identified (*Figure 3.6 panel f*). In the end, rather than analyzing four carbon dimensions in four 3D spectra, the sequential correlation is unambiguously identified with the single ^{1}H trace of HNH$_s$caco shown in *Figure 3.6 panel f.*

Optimal covariance sequential correlation maps are obtained when all dimensions of the original spectra are probed. So far, we have investigated the quality of covariance maps in a situation that exacerbates the effect of spectral crowding, namely by using a 2D projection of the 3D-HN(CO)CA. However, in practice, two three-dimensional spectra are available, and the sequential assignment procedure can be reformulated as "find which (H,N) correlations in HN(CO)CA possess C$^\alpha$ frequencies matching those observed for (H,N) correlations in HNCA." This sentence translates into:

$$HNH_sN_sca(a,b,c,e) = \frac{1}{D-1}\sum_{d=1}^{D} HNC\widetilde{A(a},b,d)HN(C\widetilde{O)CA}(c,e,d), \quad (4)$$

The resultant 4D spectrum is the HNH$_s$N$_s$ca featuring correlations (H(*i*), N(*i*), H(*i*+1), N(*i*+1)). The index *e* spans the HN(CO)CA nitrogen dimension. However, the computational implementation of equation 4 is problematic since the 4D spectrum rapidly exceeds memory capacities. Thus, instead, the four different 3D projections of the 4D spectrum are calculated on the fly, resulting in computational time and disk-space savings. Here, our MATLAB[137] processing script (available upon request) produces four 3D COSCOMs: HNH_scaco providing (H(*i*), N(*i*), H(*i*+1)), HNN_scaco providing (H(*i*), N(*i*), N(*i*+1)), H_sN_sHcaco providing (H(*i*), N(*i*), H(*i*-1)), and H_sN_sNcaco providing (H(*i*), N(*i*), N(*i*-1)). Later, others in our group updated the MATLAB processing script to make generation of 4D COSCOMs feasible[138].

The advantages of sequential covariance spectra over traditional methods are exemplified with the 37 kDa monomeric E_A for which backbone assignment had been in progress for about 5 months. *Figure 3.7* showcases both the ease of use of COSCOMs and their ability to overcome the limitations of peak picking. Four COSCOMs were used to scan the unassigned HN-TROSY-HSQC of the protein. The backbone signals of L189-Q196 were simultaneously picked and assigned within only 30 minutes (*Figure 3.8*). HNCA, HN(CA)CO and HN(CA)CB were used for residue type assignment. In contrast, only A194, G195, and Q196 had been assigned with strip matching. Several mistakes had impeded proper assignment of this segment of residues. First, the signals of A190 were erroneously assigned to A194 as all ^{13}C sequential correlations in G195 (C^{α}, C^{β}, and CO) had frequencies matching those of A190. Second, A194 had not been identified by strip-matching because its C^{α} had been mis-assigned. Finally, the signals of L189 had not been picked. When scanning G195 with HH_sN_scaco and H_sN_sNcaco (cyan strips, *Figure 3.7 panel a* top and left), A194 (labeled with a red "+" in *Figure 3.7*) and A190 (unlabeled) were both identified. A NOESY-HN-TROSY delineated between the signals of A190 and A194.

The chain of sequential residues was rapidly extended with COSCOMs down to A190, previously erroneously assigned to A194. Weak correlations in HH_sN_scaco and H_sN_sNcaco identified a new (H,N) correlation for the predecessor of A190, L189. Most likely, L189 had previously escaped peak picking because its (H,N) correlation overlaps partially with that of a very intense signal. In the end, the complete sequence of residues L189-Q196 was assigned with COSCOMs in a matter of minutes by simple scanning of an HN-TROSY, whereas strip matching only provided the correct assignment for two of these residues.

3. Equilibrium refolding of deuterated protein

Upon inspection of *Figure 3.3*, we noticed that two stretches of residues could not be assigned despite using TROSY, preparing partially deuterated samples, and employing NUS and istHMS reconstruction. We wondered whether slow back exchange of amides from N-D (since samples were grown in 70% ^2H_2O) to N-H could be an explanation since all of the triple resonance experiments applied hitherto rely upon direct detection of H^N signals. To test this hypothesis, ^{15}N protein was grown in H_2O and HN-TROSY-HSQC spectra were acquired; then samples were lyophilized and resuspended in ^2H_2O. HN-TROSY-HSQC spectra were acquired intermittently for up to three weeks (*Figure 3.9*). At the end of three weeks, we observed that about 15 signals remain, indicating slow solvent back exchange. Given this result, we wondered whether we could unfold our purified NMR sample, then refold it (all using H_2O with no ^2H_2O as solvent), in order to facilitate amide exchange.

50 mL E_A (~ 0.5 mg mL^{-1}) was denatured using 20 mM sodium phosphate 0.2 M NaCl 10 mM DTT prepared in 8M deionized urea (BioRad Mixed Bed Resin), pH 6.8 at 4˚C, and incubated for 1 hour on ice. Following denaturation, protein folding was achieved by serial dialysis for at least 4 steps of dialysis against 4L of 20 mM sodium phosphate 0.2 M NaCl 10 mM DTT pH 6.8 for at least 2 hours (at 4˚C). Finally, dialysate was concentrated to ~ 2.0 mL and injected on a Superdex 75 16/60 column (GE Healthcare) equilibrated in 20 mM Hepes 0.2 M NaCl 10 mM DTT pH 6.8. "Refolded" E_A eluted at the same position as native- E_A sample indicating refolding success. In order to confirm that the unfolding-refolding transition occurred, fluorescence spectra of samples were acquired. Using an excitation wavelength of 280 nm, fluorescence emission spectra were compared of native E_A, unfolded E_A (in buffered urea), and refolded E_A (from size exclusion chromatography) (*Figure 3.10*). The characteristic bathochromatic shift (red shift) of the tryptophan fluorescence emission reflects that the protein has completely unfolded in the presence of 8 M urea. Encouragingly, there is no detectable change (by fluorescence emission spectroscopy) from native E_A and refolded E_A; therefore, we conclude that the protein has refolded to its native state. Further confirmation that the protein came to its native state came form NMR spectroscopy. HN-TROSY-HSQC spectra were taken of E_A and refolded E_A. A 2D overlay is presented in *Figure 3.11*, which shows minor

CSPs in the refolded sample, that some new signals appear, and that others have a dramatic increase in sensitivity.

4. 80% of backbone resonance assignment & impediments to further assignment

70-DCN E_A was prepared and subjected to equilibrium unfolding/refolding as before. The same suite of triple resonance experiments were collected, including (now this time) the TROSY-HN(COCA)CB, otherwise thought too insensitive (rescued by collection of 144 transients, with 5% sampling). (See *Table 3.2* for NMR data acquisition parameters.) 4D COSCOMs were generated using these data. Using the strategies designed in this chapter, backbone assignment was brought to nearly 80% of completion. Using the available C^{α} and C^{β} chemical shift calculations, we ran a calculation with the TALOSN server[118] (Bax group), which uses C^{α} and C^{β} chemical shift information to calculate backbone dihedral angles. The calculated dihedral values are in good agreement with the PSI-PRED predictions, which gives us confidence in our assignment. (*Figure 3.12*) The ^{13}C-detected experiment, HACACO, was collected on refolded E_A and allowed for the assignment of 201 H^{α} signals.

While analyzing data from the second-round of data collection (that is, 70-DCN refolded E_A), we realized that refolded E_A undergoes some amount of degradation. Consequently, some peaks are “doubled” reflecting two chemical environments (native and degraded). (*Figure 3.13*) This level of degradation evaded our standard protein characterization (size exclusion chromatography and SDS-PAGE analysis). Indeed, it was only upon comparison of triple resonance spectra between native E_A and refolded E_A that we realized that several of the new peaks that we had previously concluded were products of solvent back exchange were actually signals from degradation products generated during our refolding protocol. Attempts to prevent proteolysis by supplementing NMR samples with EDTA and Roche Complete Protease Inhibitor Tablets (Roche) resulted in sample instability. At this point, a decision was made to focus on the experiment in *Chapter 4*.

5. Side-chain/Methyl Assignment

As mentioned previously (*Chapter 1, section 4.iii*), it is becoming commonplace to assign the methyl resonances of large protein using samples that are U-$^{13}C^{15}N^{2}H$ and Me-ILV-$^{1}H^{13}C$[41,139,140]. Typically, experiments that take advantage of COSY steps are used to transfer magnetization as follows:

$$H^{M} \xrightarrow{\text{INEPT}} C^{M} \xrightarrow{\text{COSY steps}} C^{\alpha} + C^{\beta}(t_1) \xrightarrow{\text{COSY steps}} C^{M}(t_2) \xrightarrow{\text{INEPT}} H^{M}(t_3)$$

After INEPT transfer of polarization form H^M to C^M, a series of COSY steps, allows for transfer along the carbons of the side chains of methyl-bearing amino acids. The use of COSY transfers was initially obvious to Tugarinov and Kay[52,53] since (i) complete transfer of in-phase magnetization to anti-phase magnetization (or the opposite) is possible in 14 ms and (ii) transfers are directional such that the coherence transfer trajectory pushes magnetization unidirectionally from C^M to C^α. Alternatives to this strategy would include ^{13}C-^{13}C NOESY or ^{13}C-^{13}C TOCSY-type experiments. Unlike COSY transfers, TOCSY transfers are not unidirectional as magnetization is mixed (in an optimal scenario) evenly along the side chain. Secondly, optimal TOCSY mixing requires long TOCSY loops (on the order of 30 ms), during which coherences are subjected to disastrous relaxation rates.

Our previous use of COSY-type transfers on a 53 kDa Cy1 domain (the sample used exclusively in *Chapter 4*) exhibited good sensitivity, but spectra were plagued with multiple quantum artifacts that are an unavoidable byproduct of sequences featuring multiple homonuclear COSY steps. Assignment of ILV resonances from these COSY-type experiments requires using $\{H^M, C^M\}$ spin systems as anchors. For Leu and Ile, $\{C^\gamma\}$ is connected to the $\{H^M, C^M\}$ anchor; for Ile, Leu, Val, $\{C^\alpha$ and $C^\beta\}$ are connected to the $\{H^M, C^M\}$ anchor. Finally, C^α and C^β are matched to $\{H^N, N\}$ anchors, allowing for the $\{H^M, C^M\}$ spin systems to be merged with the $\{H^N, N\}$ spin systems. In an effort to avoid these indirect assignment methods (that require working through both $\{H^M, C^M\}$ anchors and $\{H^N, N\}$ anchors) and to acquire spectra without multiple quantum artifacts, we decided to adapt the well-known H(CCCCO)NH and (H)C(CCCO)NH experiments for U-$^{13}C^{15}N^2H$ and Me-ILV-$^1H^{13}C$ samples and to use NUS to rescue sensitivity. The magnetization trajectory for the TROSY-H(CCCCO)NH is shown first below, and TROSY-(H)C(CCCO)NH is shown second below (applied to methyl-bearing amino acids):

$$H_i^M(t_1) \xrightarrow{\text{INEPT}} C_i^M \xrightarrow{\text{TOCSY}} C_i^M + C_i^\beta + C_i^\alpha \xrightarrow{\text{INEPT}} N_{i-1}(t_2) \xrightarrow{\text{TROSY}} H^N_{i-1}(t_3)$$

$$H_i^M \xrightarrow{\text{INEPT}} C_i^M(t_1) \xrightarrow{\text{TOCSY}} C_i^M + C_i^\beta + C_i^\alpha \xrightarrow{\text{INEPT}} N_{i-1}(t_2) \xrightarrow{\text{TROSY}} H^N_{i-1}(t_3)$$

$H^M_x(i)$ is transferred to $2H^M_y C^M_z$ (i) using INEPT transfers. A refocused spin-echo, converts the coherence into C^M_x, and a series of long TOCSY pulse trains distributes magnetization along the carbon side chain. After TOCSY, two refocused spin echoes are used to first convert C^{α}_x into $2C^{\alpha}_yCO_z$ then to convert CO_y into $2N_zCO_x$. Finally, as in TROSY-HNCO, back transfer to ^{15}N typically occurs in a constant time manner, and TROSY line selection can be used to select for the most slowly relaxing component of the ^{1}H^{15}N spin system. Even when non-methyl aliphatic carbons are perdeuterated, these sequences do not generally deliver sufficient sensitivity for larger proteins.

Since we were previously successful using NUS to rescue sensitivity of otherwise insensitive experiments (most encouragingly, the HN(COCA)CB, *Section 4*), we wondered whether we could apply NUS to allow for a high number of scans on the TROSY-H(CCCCO)NH and TROSY-(H)C(CCCO)NH. In order to optimize these experiments for methyl groups, a few modifications were made to the standard pulse sequences. First, we placed our ^{13}C carrier frequently on-resonance with methyl signals (19 ppm or 2850 Hz at 14.1 T). Secondly and as with other experiments, we employed concatenation in the back-transfer pathway of the $C^{\alpha} \rightarrow N$, as described in *Section 1*), and encoded the nitrogen chemical shift in a semi-constant time (*Section 1*) period. Using NUS, we were able to collect 256 scans of the TROSY-H(CCCO)NH and 224 scans of the TROSY-(H)C(CCO)NH by sampling only 10% of indirect dimension points for both experiments. 2D projections of the 3D data sets are shown in *Figure 3.14*. Inspection of the 2D projections reveals sensitivity beyond our expectations for both experiments. We underscore that these data are among the few TROSY-H(CCCO)NH or TROSY-(H)C(CCO)NH data sets that, to our knowledge, have been collected on a monomeric protein of this size. Using these two experiments, we were able to directly connect $H^M(i)$ and $C^M(i)$ to $\{H^N(i+1), N(i+1)\}$ spin systems. Of the 78 methyl-bearing residues (9 Ile, 56 Leu, and 13 Val), 3 cannot be assigned because they precede proline residues (not detectable signals in TROSY-H(CCCO)NH or TROSY-(H)C(CCO)NH), and another 11 cannot be assigned because their successors have not yet been assigned; there are currently methyl assignment for 31 (3 Ile, 9 Leu, and 19 Val) of the remaining 64 assignable residues. (*Figure 3.15*)

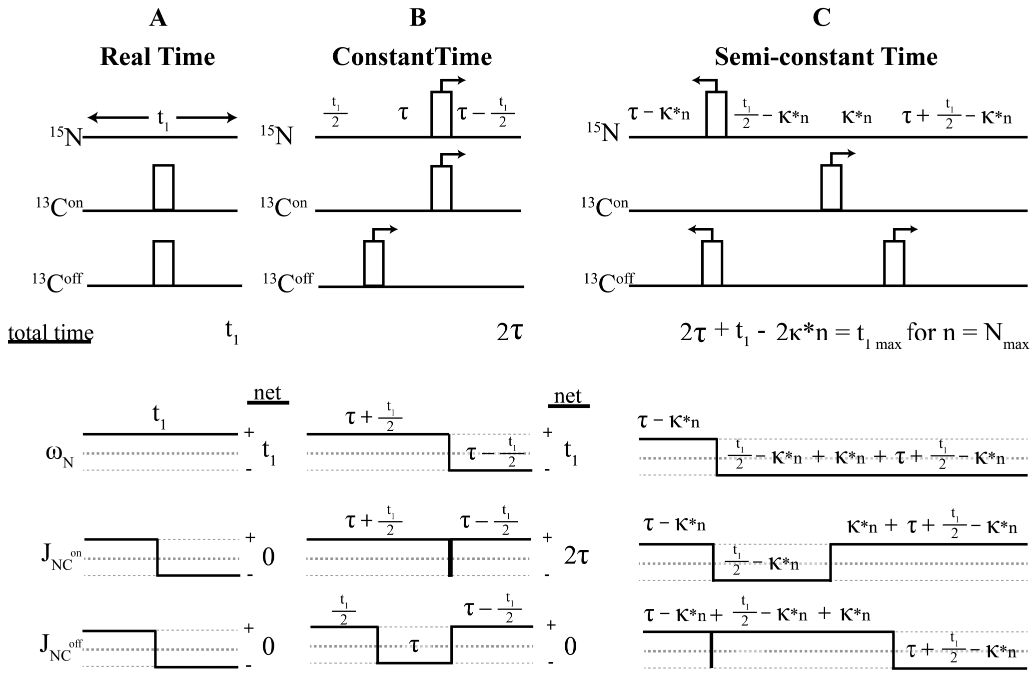


Figure 3.1 Comparison of real time, constant time, and semi-constant time pulse sequence blocks. J_{NCon} is the coupling constant between ^{15}N and $^{13}C_{on}$, where $^{13}C_{on}$ is the on resonance nucleus (e.g. ^{13}CO for HNCO), and J_{NCoff} is the coupling constant between ^{15}N and $^{13}C_{off}$ (e.g. $^{13}C^{\alpha}$ for HNCO). (a) In a real time encoding period, the density matrix operator evolves only under ^{15}N chemical shift. (b) In this constant time encoding period, the density matrix operator evolves under ^{15}N chemical shift as a function of t_1 and under J_{NC} for 2τ, where $2\tau = 1/2J_{NCon}$, allowing for in-phase to anti-phase magnetization transfer to occur. Evolution is refocused under J_{NCoff}. (c) This semi-constant time encoding period is similar to the constant time encoding period, as it allows for the same evolutions of the density matrix operator; however, larger resolutions can be achieved. $\kappa = \tau / N_{max}$, where $\tau = 1/4J_{NCon}$, and N_{max} = total number of indirect dimension points; n = increment number.

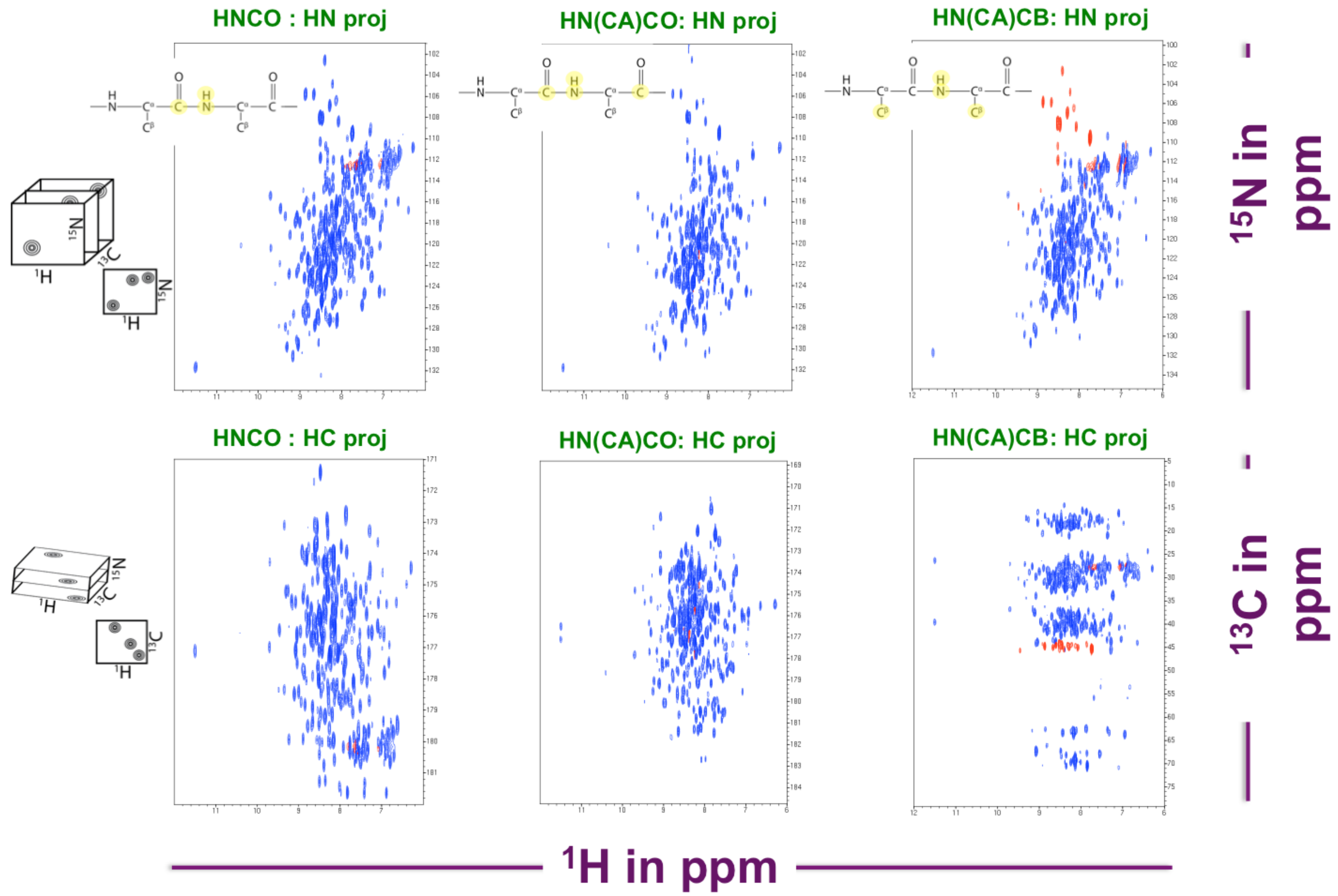
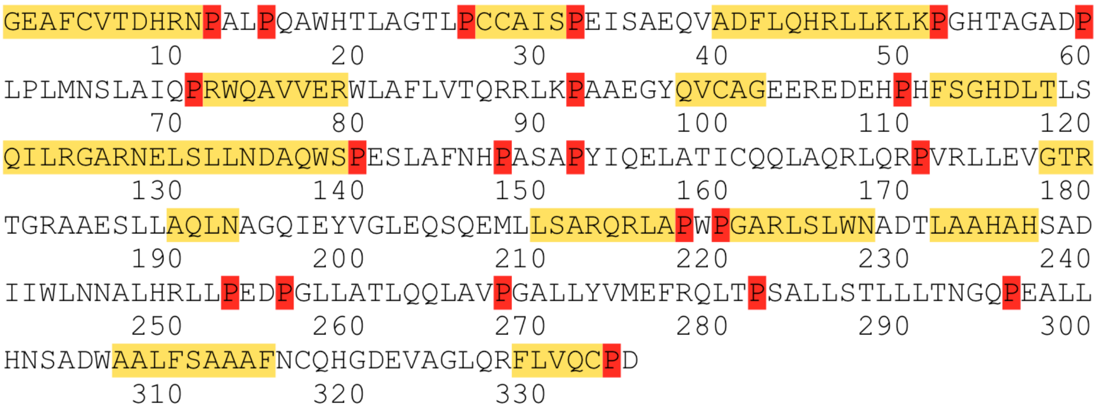

HNCO : HN proj
HN(CA)CO: HN proj
HN(CA)CB: HN proj
HNCO : HC proj
HN(CA)CO: HC proj
HN(CA)CB: HC proj
^{15}N in ppm
^{13}C in ppm
^{1}H in ppm
^{1}H
^{15}N
^{13}C

Figure 3.2 Representative spectra collected on 70-DCN E_A. Projections (H/N, *top* and H/C, *bottom*) of triple resonance spectra acquired on $^{15}N^{13}C$-E_A grown in 70% ^2H_2O/30% ^1H_2O and recorded on a 600 MHz at 298 K in a 400 μL sample (prepared in a Shigemi tube). Refer to table 3.1 for more information on NMR data collection. *Left*: projections of the HNCO, most sensitive of triple resonance experiments that generates correlations of the form (H^N_i, N_i, CO_{i-1}). *Center*: projections of the HN(CA)CO, moderately sensitive triple resonance experiment that generates correlations of the form (H^N_i, N_i, CO_i) & (H^N_i, N_i, CO_{i-1}). *Right:* projections of the HN(CA)CB, the least sensitive of these triple resonance experiments that generates correlations of the form (H^N_i, N_i, C^β_i) & (H^N_i, N_i, C^β_{i-1}).



```
GEAFCVTDHRNPALPQAWHTLAGTLPCCAISPEISAEQVADFLQHRLLKLKPGHTAGADP
        10        20        30        40        50        60
LPLMNSLAIQPRWQAVVERWLAFLVTQRRLKPAAEGYQVCAGEEREDEHPHFSGHDLTLS
        70        80        90       100       110       120
QILRGARNELSLLNDAQWSPESLAFNHPASAPYIQELATICQQLAQRLQRPVRLLEVGTR
       130       140       150       160       170       180
TGRAAESLLAQLNAGQIEYVGLEQSQEMLLSARQRLAPWPGARLSLWNADTLAAHAHSAD
       190       200       210       220       230       240
IIWLNNALHRLLPEDPGLLATLQQLAVPGALLYVMEFRQLTPSALLSTLLLTNGQPEALL
       250       260       270       280       290       300
HNSADWAALFSAAAFNCQHGDEVAGLQRFLVQCPD
       310       320       330
```

Figure 3.3 Backbone assignment progress. Single letter amino acid code of the primary sequence of E_A is shown. Residues highlighted in yellow were assigned using only low resolution triple resonance data and strip-matching procedures to assign backbone resonances; approximately 30% of the backbone of E_A was assigned. Prolines, which do not contain H^N are depicted in red since they cannot be assigned with the available data.

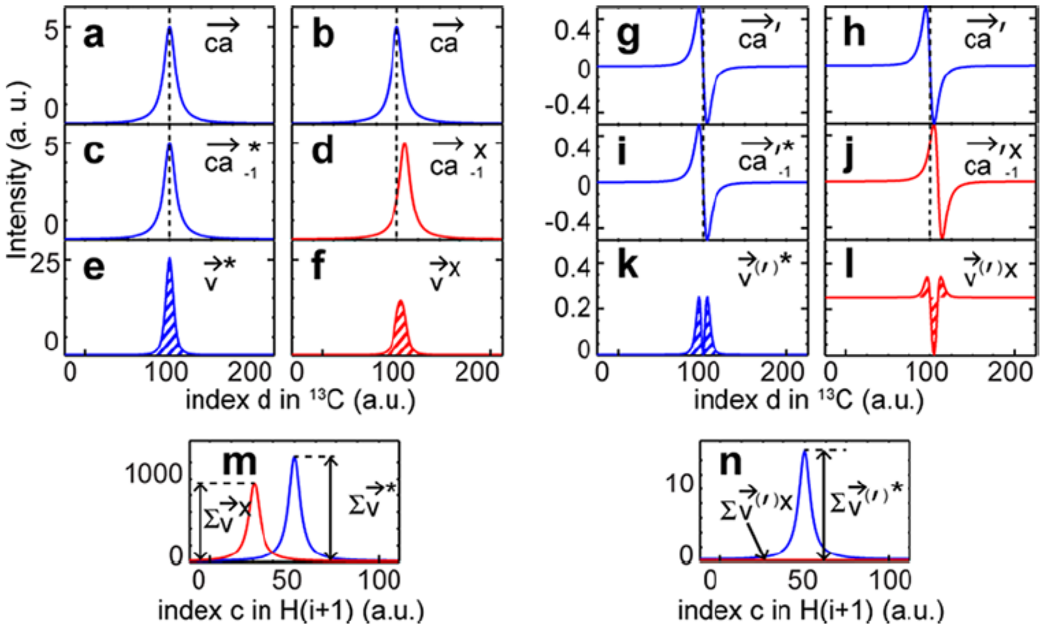


Figure 3.4 Spectral derivatives suppress spurious correlations in covariance NMR spectra. The * and *X* indicate the true and erroneous correlations, respectively; (a,b) $\overrightarrow{ca}$ at an index H(*i*) = *a* (see equations 2 and 3); (c) $\overrightarrow{ca}^{*}_{-1}$at an index H(*i*+1) = c^{*}; (d) $\overrightarrow{ca}^{X}_{-1}$for an erroneous correlation at H(*i*+1) = c^{X}. (e,f) Element-wise products of $\overrightarrow{ca}$ with $\overrightarrow{ca}^{*}_{-1}$ ($\vec{v}^{X}$). (g—j) Derivatives ($\overrightarrow{ca}'$) of $\overrightarrow{ca}$ vectors in a—d, respectively. (k,l) Element-wise products of $\overrightarrow{ca}'$ with $\overrightarrow{ca}'^{*}_{-1}$ ($\vec{v}^{(\prime)X}$). $\vec{v}^{(\prime)*}$ and $\vec{v}^{(\prime)X}$ denote the products of the derivatives and not the derivatives of the products. (m) H(*i*+1) trace in $\mathrm{HH_s}$ at index H(*i*) = *a*, *without* derivatives. (n) Corresponding H(*i*+1) trace *with* derivatives.

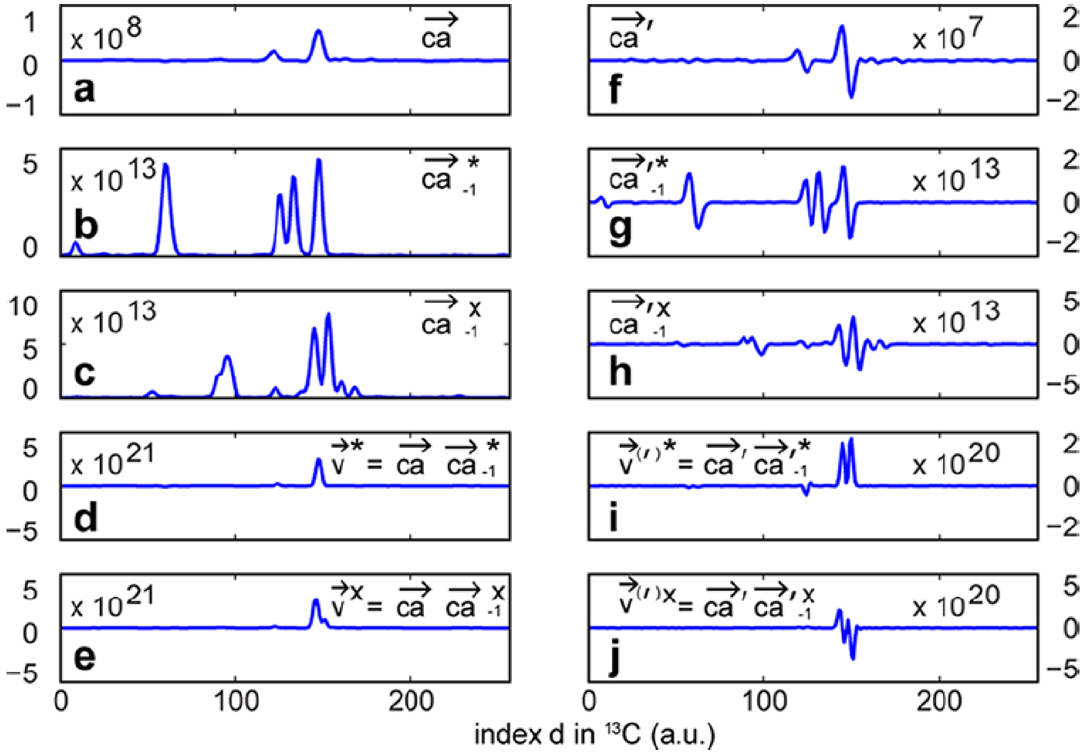


Figure 3.5 Differentiating between true (*) sequential and erroneous correlations due to partially overlapping signals (X). (a) $\overrightarrow{ca}$ (C$^{\alpha}$ 1D trace) from HNCA at ^{1}H(i) = 7.558 ppm and ^{15}N(i) = 120.023 ppm; (b) $\overrightarrow{ca_{-1}}^{*}$ from 2D-H(NCO)CA at ^{1}H(i+1) = 7.608 ppm; (c) $\overrightarrow{ca_{-1}}^{X}$ from 2D-H(NCO)CA at ^{1}H(i+1) = 8.602 ppm. (d) Element-wise product of $\overrightarrow{ca}$ with $\overrightarrow{ca_{-1}}^{*}$ ($\vec{v}^{*}$). (e) element-wise product of the traces in a and c. f-h) derivatives of $\overrightarrow{ca}$ vectors in a-c, respectively. i-j) element-wise products of $\overrightarrow{ca}$ with $\overrightarrow{ca_{-1}}^{X}$ ($\vec{v}^{X}$). (f—h) Derivatives of $\overrightarrow{ca}'$ vectors in a—c, respectively. (i,j) Element-wise products of $\overrightarrow{ca}'$ with $\overrightarrow{ca}'^{*}_{-1}$ ($\vec{v}^{(\prime)*}$) and $\overrightarrow{ca}'$ with $\overrightarrow{ca}'^{X}_{-1}$ ($\vec{v}^{(\prime)X}$), respectively. The normalized sum of the elements $\vec{v}^{*}$, $\vec{v}^{(\prime)*}$, $\vec{v}^{X}$, and $\vec{v}^{(\prime)X}$ lead to correlations that are highlighted by the symbols * and X in *Figure 3.5*, respectively.

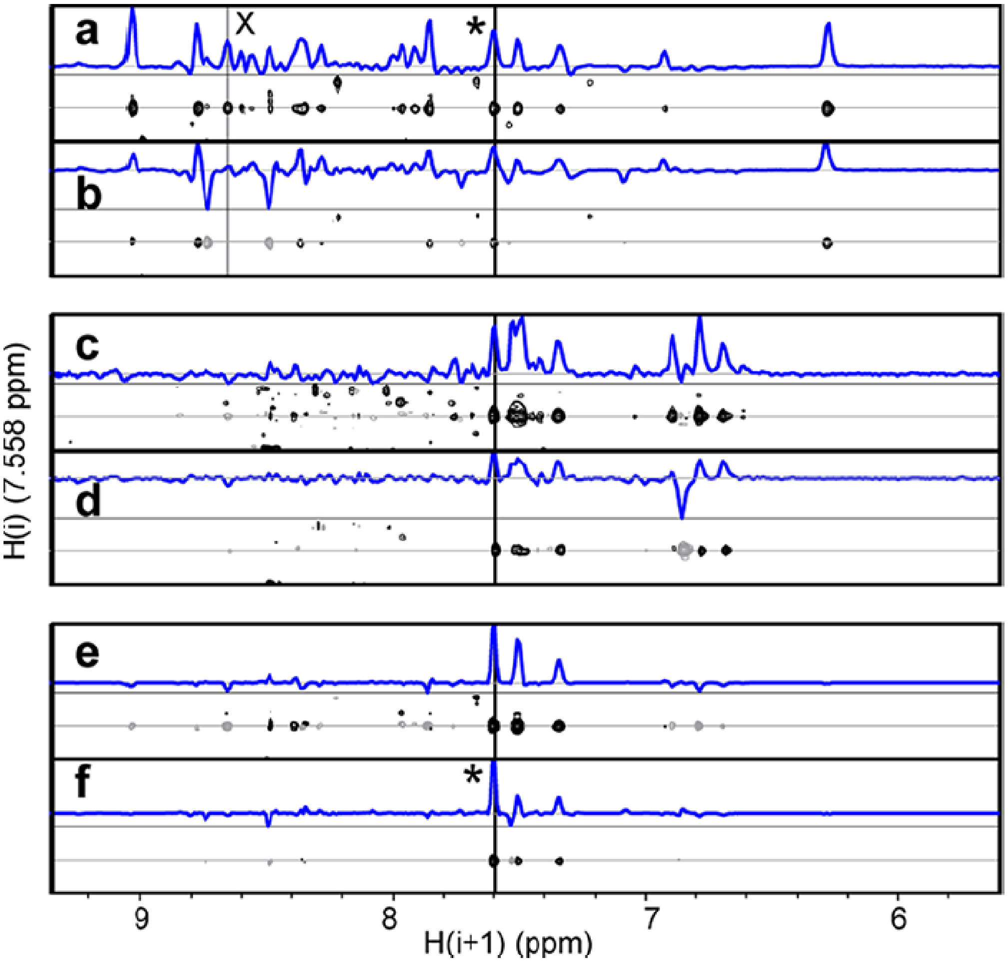


Figure 3.6 Identification of unique proton sequential correlations when using spectral derivatives and when multiplying COSCOMs (covariance sequential correlation maps). (a,b) $\mathrm{HNH_S}$*ca*, (c,d) $\mathrm{HNH_S}$*co*, and (e,f) $\mathrm{HNH_S}$*caco* obtained by multiplying *a* and *c* and *b* and *d*, respectively. (a,c,e) Correlations obtained without derivatives in the carbon dimensions. (b,d,f) Correlations obtained with derivatives. Covariance was performed with the MATLAB[137] covariance NMR toolbox[130]. The amplitudes of signals labeled * are $\sum \vec{v}^{*}$ and $\sum \vec{v}^{(\prime)*}$ in a and b, respectively, while those labeled *X* are $\sum \vec{v}^{X}$ and $\sum \vec{v}^{(\prime)*X}$, with the vectors $\vec{v}$as defined in *Figure 3.4*. The * denotes the true correlation.

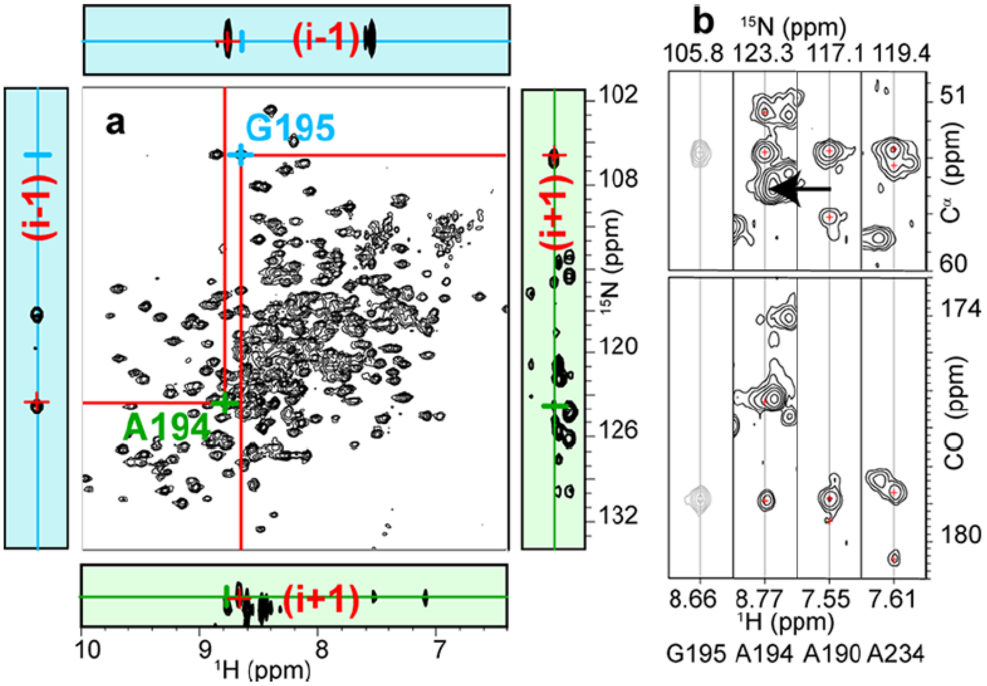


Figure 3.7 Scanning HN-TROSY-HSQC of COSCOMs (covariance sequential correlation maps) overcomes shortfalls of strip matching. (a) HN-TROSY-HSQC of E_A with strips of HNH_{pre} (left) and HNN_{pre} (top) at the (H,N) coordinates of G195 (cyan), as well s strips of HNH_{suc} (bottom) and HNN_{suc} (right) at the coordinates of A194 (green). (b) Strip matching for the predecessor of G195. A194 was initially missing; its C^{α} was erroneously picked at the position indicated by the arrow. Correlations to A190 and A234 (very weak) are seen in HNH_{pre} and HNN_{pre} (unlabeled).

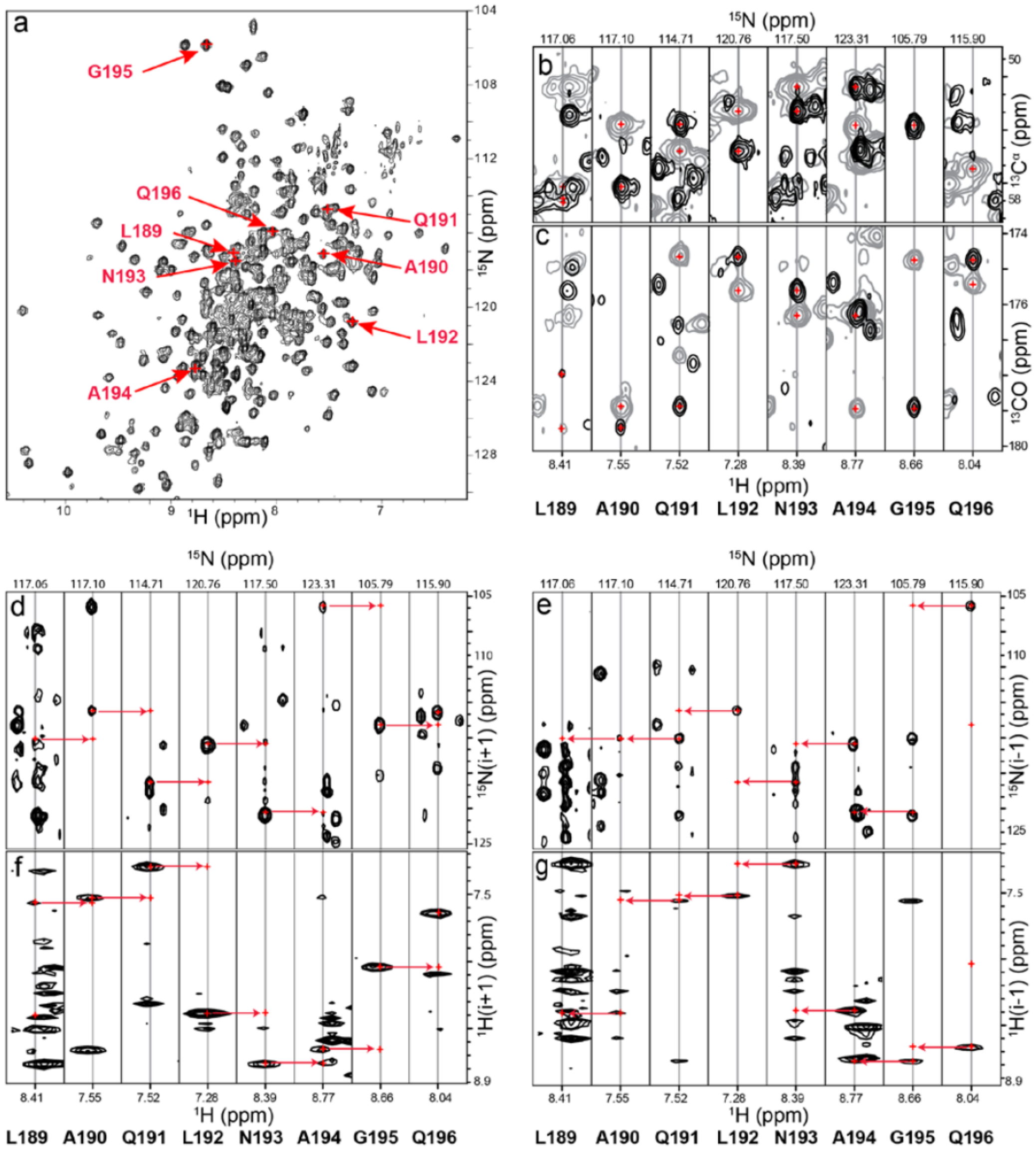


Figure 3.8 Comparison of the assignment performed with strip matching and COSCOMs (covariance sequential correlation maps). (a) HN-TROSY-HSQC of 70%-^{2}H-U-^{15}N-^{13}C 37 kDa EA. Signals of residues that have been assigned by COSCOMs during this comparison are labeled in red. (b) HN(CO)CA (black) and HNCA (grey) strips for the segment L189-Q196. (c) HNCO (black) and HN(CA)CO (grey). Among these residues, only G195 and A196 were assigned correctly with strip-matching. (d,e,f,g) strips of 3D projections of HNH$_{s}$N$_{s}$. The horizontal axis in

each strip is along the non-sequential proton dimension. (d) HNN_s provides N(*i*+1), (e) H_sN_sN provides N(*i*-1), (f) HNH_s provides H(*i*+1), and (g) H_sN_sH provides H(*i*-1). The names of the spectra as provided by our script are HNN_{suc}, HNN_{pre}, HNH_{suc}, and HNH_{pre} for *d, e, f,* and *g*, respectively. Red arrows highlight sequential correlations and define the direction of chain elongation.

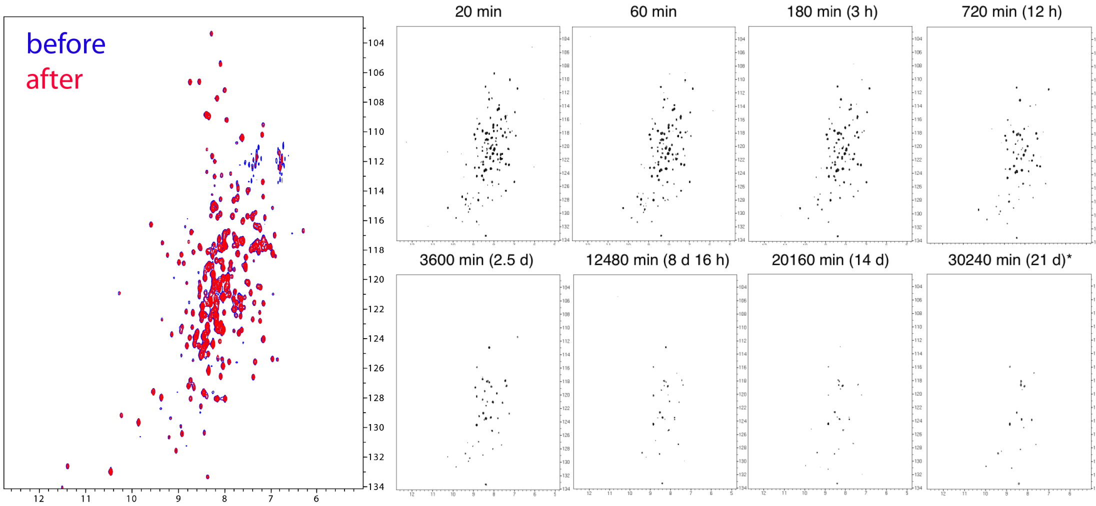


Figure 3.9 Hydrogen-deuterium exchange data as a function of time. 600 µL samples of $^{1}H^{15}N$-E_A (900 µM) were purified in 20 mM Hepes, 0.2 M NaCl, 10 mM DTT, pH 6.8, lyophilized and resuspended in $^{2}H_2O$. Data were collected at 600 MHz at 298 K. 512 and 128 complex points were collected in the ^{1}H and ^{15}N dimensions, respectively. Eight transients were collected for each spectrum, with a one second recycling delay. On the left, samples before and after lyophilization are shown in blue and red, respectively. HN-TROSY-HSQC spectra were taken for up to 21 days after resuspension and are shown on the right. We observed that even after ~ 14 days (total acquisition time of data used hitherto for backbone assignment) approximately 15-20 residues have still not back exchanged from N-H to N-D.

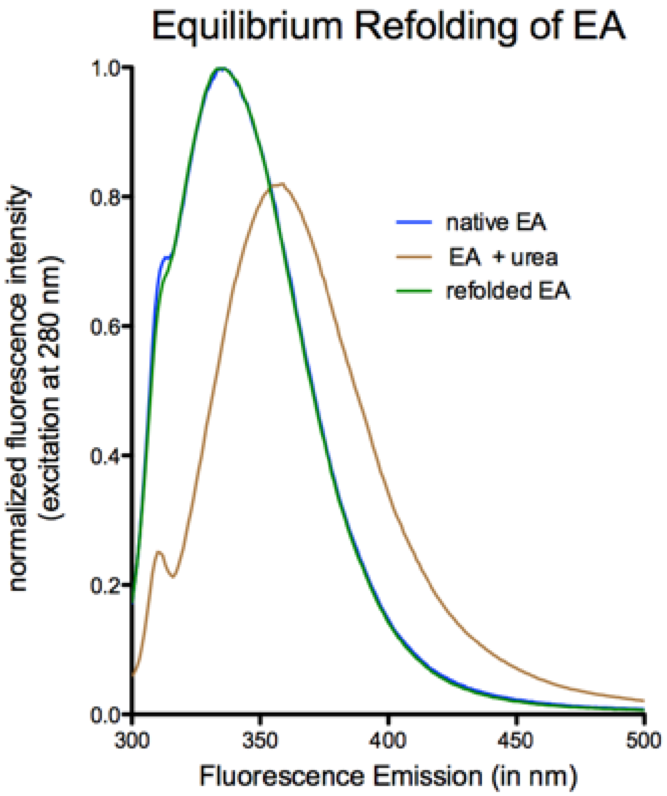


Figure 3.10 Demonstration of equilibrium refolding of E_A by tryptophan fluorescence emission spectroscopy. Native fluorescence emission spectrum of tryptophan (excitation at 280 nm) of native E_A (blue), unfolded E_A in 8 M urea (brown), and refolded E_A (green) was monitored from 300 nm – 500 nm. Native EA is completely unfolded by 8 M urea as judged by the large (~ 20 nm) bathochromatic shift, indicating fluorescent tryptophans are fully exposed to polar solvent (water). The native E_A and refolded E_A spectra overlay nearly perfectly reflecting that the protein has globally refolded to its native state. The peak at 310 nm is the characteristic Raman scattering signal in water (for excitation at 280 nm).

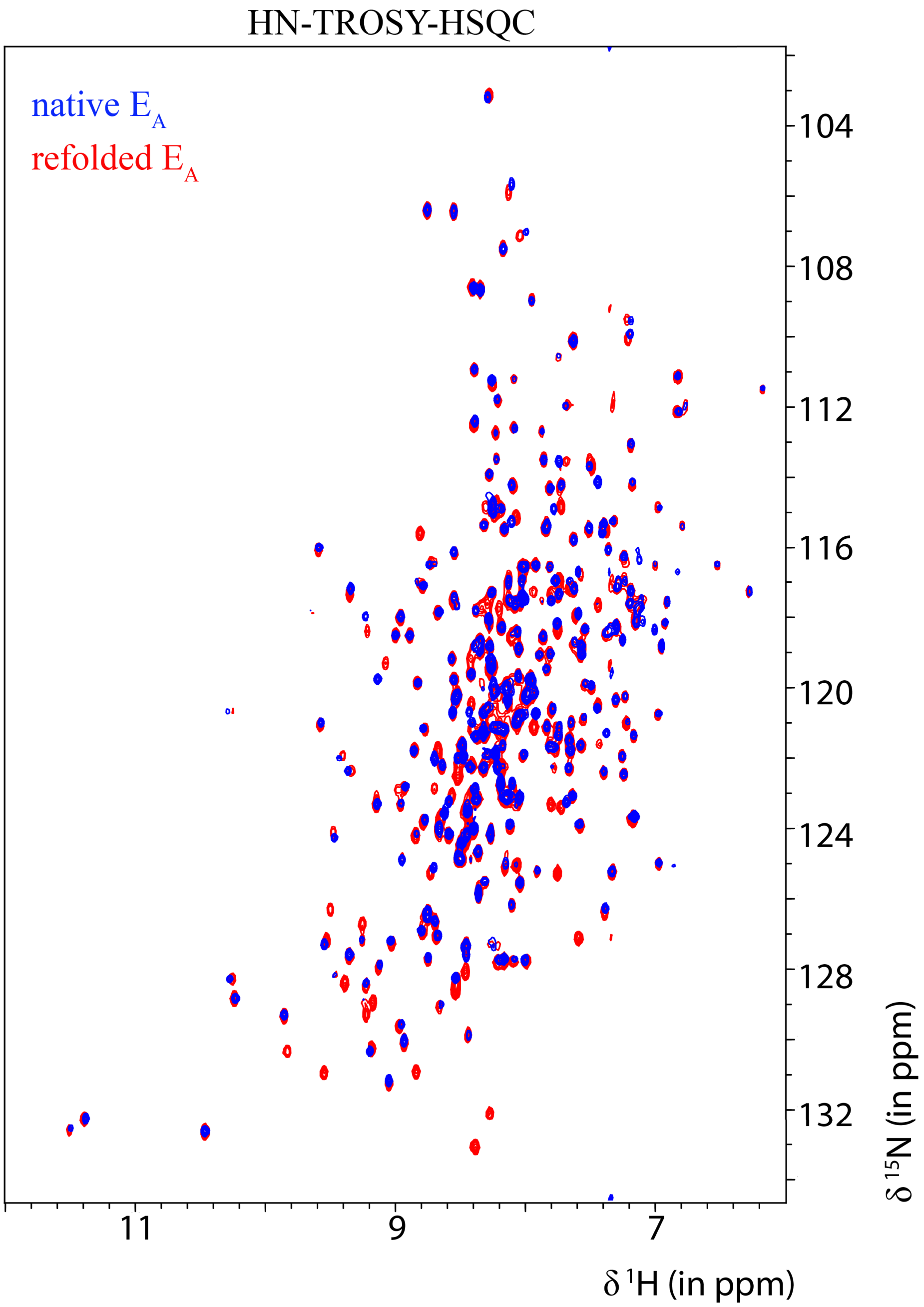


Figure 3.11 Overlay of HN-TROSY-HSQC of native E_A and refolded E_A. HN-TROSY-HSQC spectra were recorded on 600 μM 70-DCN native E_A and then on refolded 600 μM refolded E_A. Since sample concentrations were equal, spectra are shown at the same contour levels. Samples were purified in 20 mM Hepes, 0.2 M NaCl, 10 mM DTT, pH 6.8, 5% ^2H_2O (for lock). 512 and 128 complex points were acquired in the ^1H and ^{15}N dimensions, respectively on a 600 MHz spectrometer at

298 K. For each spectrum, 8 transients were recorded with a 1 second recycling delay. Inspection on HN-TROSY-HSQC revealed that few peaks had shifted and that refolded E_A showed more peaks than native E_A, originally leading us to conclude that N-D had back-exchanged to N-H during refolding in ^1H_2O solvent.

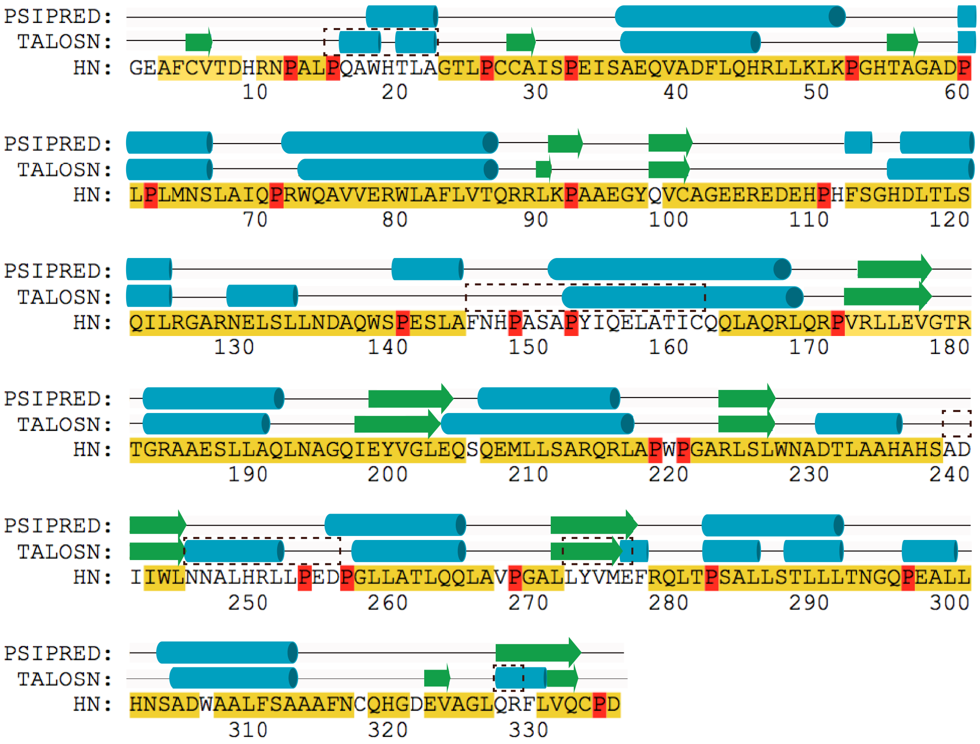


Figure 3.12 Current backbone assignment (~80%) of E$_A$. Comparable to *Figure 3.2*, single letter amino acid code of the primary sequence of E$_A$ is shown. Residues highlighted in yellow were assigned using strip matching and COSCOMs **(covariance sequential correlation maps)**; approximately 80% of the backbone of E$_A$ has been assigned. Prolines, which do not contain H^N are depicted in red since they are able to be assigned with the available data. Secondary structure predictions from the PSIPRED server[114] (top line of each set) are in good agreement with the TALOSN server[142], which uses experimentally derived chemical shift data (primarily of C$^\alpha$ and C$^\beta$ chemical shifts) to predict dihedral angles and ergo secondary structure. (For residues without chemical shift assignments, TALOSN has a routine for secondary structure estimation from primary sequence; these predictions are shown with a dashed box.)

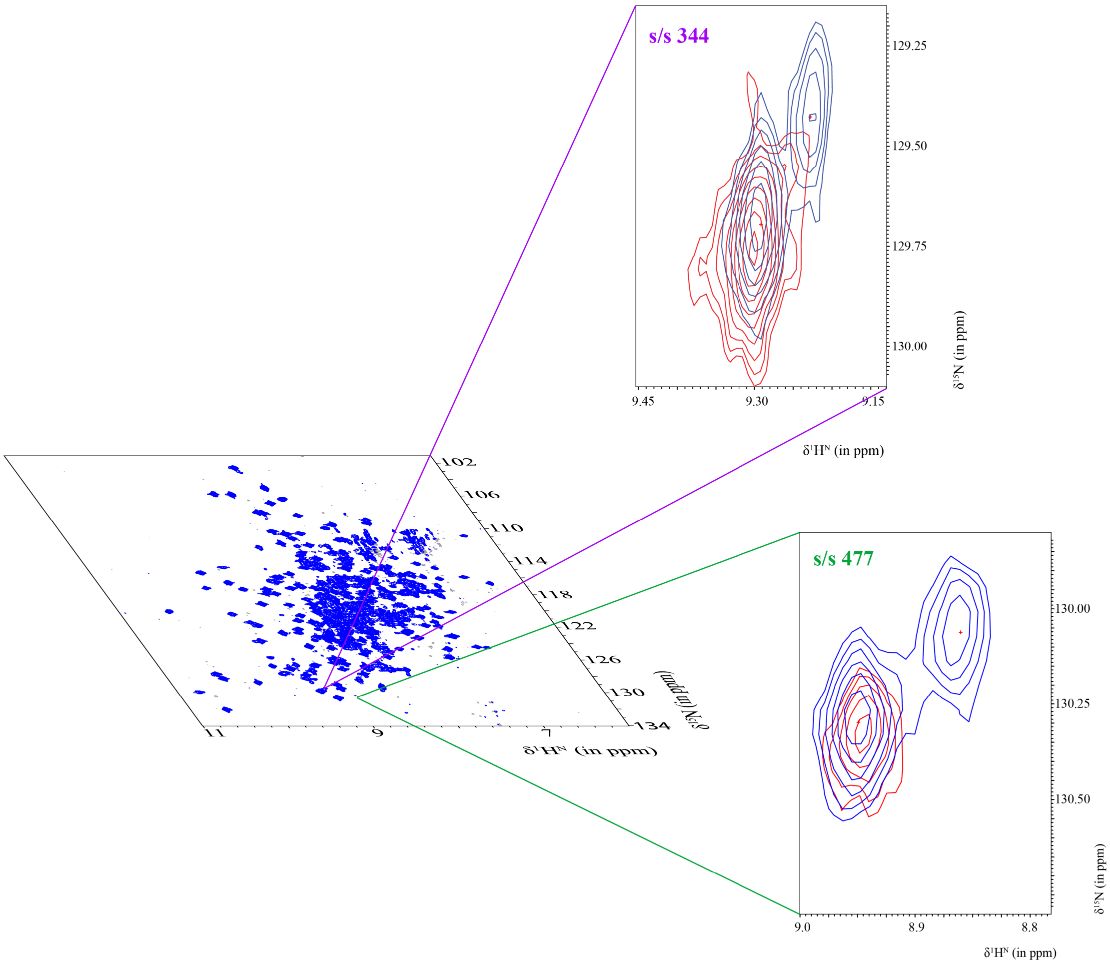


Figure 3.13 Evidence of degradation of E$_A$. Overlays of HN-TROSY-HSQC are shown of two unassigned spin systems, numbered **344 (purple)** and **477 (green).** The inlets show spectra recorded immediately after sample preparation **(red)** and two weeks after sample preparation **(blue)**; this clear peak doubling likely is a result of proteolysis that did not manifest by another laboratory technique (specifically neither by SDS-PAGE nor by size exclusion chromatography). E$_A$ was prepared in 20 mM Hepes, 0.2 M NaCl, 10 mM DTT pH 6.8, 5% ^2H_2O (for lock). Data were collected on a 600 MHz spectrometer at 298 K. Four transients were recorded with a 1 second

recycling delay. 512 and 128 complex points were recorded in the ^{1}H and ^{15}N dimensions, respectively.

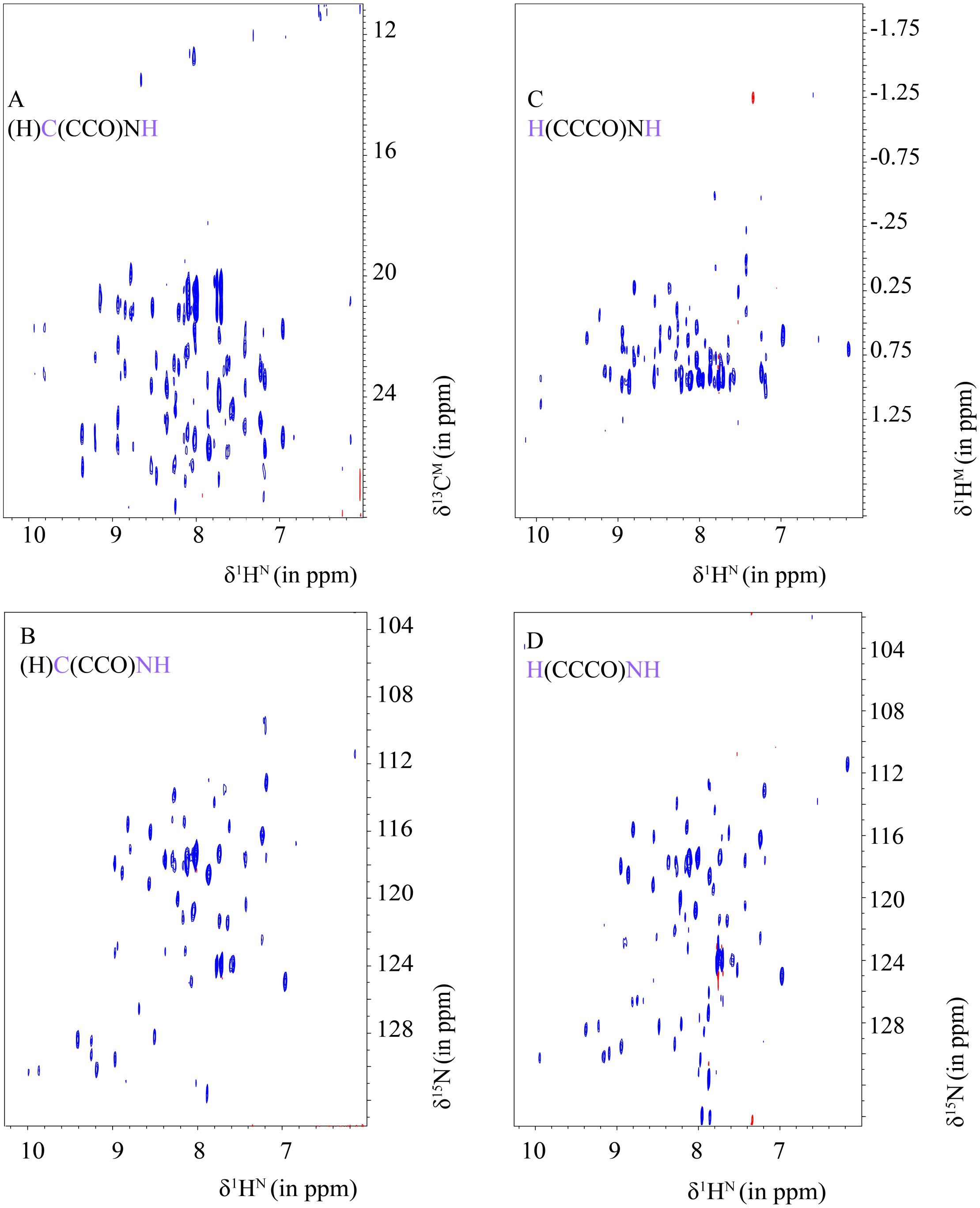
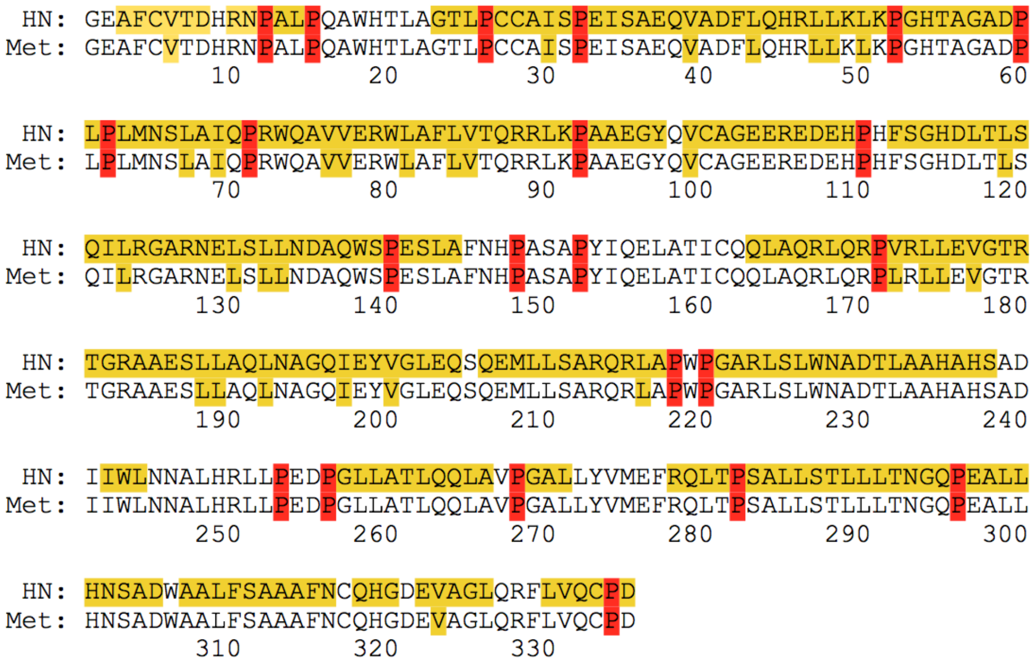


Figure 3.14 2D projections of 3D (H)C(CCO)NH and H(CCCO)NH experiments recorded on U-^{13}C^{15}N^{2}H Me-ILV-^{1}H^{13}C. Samples were prepared in 20 mM Hepes, 0.2 M NaCl, 10 mM DTT, pH 6.8, 5% $^{2}H_2O$. Spectra were collected on a 600 MHz spectrometer at 298 K. For H(CCCO)NH, 256 transients were collected (using a 1.5 second recycling delay) and 512, 45, and 45 complex points were recorded in the

direct ^{1}H, ^{15}N, and indirect ^{1}H dimensions, respectively. For (H)C(CCO)NH, 224 transients were collected (using a 1.5 second recycling delay) and 512, 45, and 45 complex points were recorded in the direct ^{1}H, ^{15}N, and ^{13}C dimensions, respectively. A 30 ms mixing time was implemented in both H(CCCO)NH and (H)C(CCO)NH. Projected dimensions are indicated in purple in the figure labels. (a) H/C projection of (H)C(CCO)NH (b) H/N projection of (H)C(CCO)NH (c) H/H projection of H(CCCO)NH (d) H/N projection of H(CCCO)NH.



```
 HN: GEAFCVTDHRNPALPQAWHTLAGTLPCCAISPEISAEQVADFLQHRLLKLKPGHTAGADP
Met: GEAFCVTDHRNPALPQAWHTLAGTLPCCAISPEISAEQVADFLQHRLLKLKPGHTAGADP
                  10        20        30        40        50        60

 HN: LPLMNSLAIQPRWQAVVERWLAFLVTQRRLKPAAEGYQVCAGEEREDEHPHFSGHDLTLS
Met: LPLMNSLAIQPRWQAVVERWLAFLVTQRRLKPAAEGYQVCAGEEREDEHPHFSGHDLTLS
                  70        80        90       100       110       120

 HN: QILRGARNELSLLNDAQWSPESLAFNHPASAPYIQELATICQQLAQRLQRPVRLLEVGTR
Met: QILRGARNELSLLNDAQWSPESLAFNHPASAPYIQELATICQQLAQRLQRPLRLLEVGTR
                 130       140       150       160       170       180

 HN: TGRAAESLLAQLNAGQIEYVGLEQSQEMLLSARQRLAPWPGARLSLWNADTLAAHAHSAD
Met: TGRAAESLLAQLNAGQIEYVGLEQSQEMLLSARQRLAPWPGARLSLWNADTLAAHAHSAD
                 190       200       210       220       230       240

 HN: IIWLNNALHRLLPEDPGLLATLQQLAVPGALLYVMEFRQLTPSALLSTLLLTNGQPEALL
Met: IIWLNNALHRLLPEDPGLLATLQQLAVPGALLYVMEFRQLTPSALLSTLLLTNGQPEALL
                 250       260       270       280       290       300

 HN: HNSADWAALFSAAAFNCQHGDEVAGLQRFLVQCPD
Met: HNSADWAALFSAAAFNCQHGDEVAGLQRFLVQCPD
                 310       320       330
```

Figure 3.15 Current backbone & methyl resonance assignment of E_A.

Comparable to *Figure 3.2*, single letter amino acid code of the primary sequence of E_A is shown. Residues highlighted in yellow on the “HN” line reflect assignment of amide ($^1H^N$, ^{15}N) resonances (and nearly complete assignment of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO). Residues highlighted in yellow on the “Met” line were assigned from TROSY-H(CCCO)NH and TROSY-(H)C(CCO)NH. Prolines, which do not contain H^N are depicted in red since they cannot be assigned with the available data.

Table 3.1 70-DCN E_A NMR acquisition parameters for backbone resonance assignment.

Experiment	Number of transients (sampling factor)	Number of points (nucleus)	Number of points (nucleus)	Number of points (nucleus)
HNCO	32 (15%)	45 (^{15}N)	50 (^{13}CO)	512 ($^{1}H^{N}$)
HN(CA)CO	128 (15%)	45 (^{15}N)	50 (^{13}CO)	512 ($^{1}H^{N}$)
HNCA	48 (15%)	45 (^{15}N)	60 ($^{13}C\alpha$)	512 ($^{1}H^{N}$)
HN(CO)CA	144 (15%)	45 (^{15}N)	60 ($^{13}C\alpha$)	512 ($^{1}H^{N}$)
HN(CA)CB	144 (15%)	45 (^{15}N)	75 ($^{13}C\beta$)	512 ($^{1}H^{N}$)

Table 3.2 70-DCN Refolded E_A NMR acquisition parameters for backbone resonance assignment

Experiment	Number of transients (sampling factor)	Number of points (nucleus)	Number of points (nucleus)	Number of points (nucleus)
HNCO	64 (5%)	45 (^{15}N)	50 (^{13}CO)	512 (^{1}H^{N})
HN(CA)CO	192 (5%)	45 (^{15}N)	50 (^{13}CO)	512 (^{1}H^{N})
HNCA	76 (5%)	45 (^{15}N)	60 (^{13}Cα)	512 (^{1}H^{N})
HN(CO)CA	144 (5%)	45 (^{15}N)	60 (^{13}Cα)	512 (^{1}H^{N})
HN(CA)CB	96 (5%)	45 (^{15}N)	75 (^{13}Cβ)	512 (^{1}H^{N})
HN(COCA)CB	144 (5%)	45 (^{15}N)	75 (^{13}Cβ)	512 (^{1}H^{N})

Chapter 4: A 3D Time-Shared HN-TROSY/HC-CT-HMQC-NOESY for acquisition of NMR distance restraints in uniformly ^{13}C enriched proteins

A portion of the following text will be submitted for publication.

1. Introduction

Nuclear Magnetic Resonance is a powerful technique for obtaining structures of proteins and nucleic acids in near physiological conditions and in the presence of dynamics and transient interactions. Due to efficient transverse relaxation (mediated by dipole-dipole interactions and chemical shift anisotropy), NMR spectroscopy suffers from signal losses as molecular weight (and also magnetic field) increases. To a large extent, these sensitivity losses are overcome upon implementation of the TROSY method (transverse relaxation optimized spectroscopy)[17] and upon replacement of aliphatic and aromatic protons with deuterons[139]. As protein deuteration renders the relaxation properties of biomolecules more favorable, it comes at the cost of decreasing the number of ^{1}H-^{1}H nOes available for structure calculations. Consequently, clever strategies have been developed in which proteins are expressed in a highly deuterated background yet methyl groups (generally of Ile, Leu, and Val) are selectively protonated using special precursors[41,62,139,143]. Protein structure determination using such a "methyl reincorporation" strategy traditionally necessitates two costly samples: In the first, all amino acids are uniformly labeled with ^{13}C and ^{15}N and exclusively protonated at methyl groups, which allows for assignment of $^{1}H^{methyl}$ and $^{13}C^{methyl}$ signals; in a second sample, all amino acids are uniformly labeled with ^{12}C and ^{15}N, and only methyl groups are labeled with ^{13}C and ^{1}H, which allows for collection of nOe distance constraints for structure determination ($H^m \rightarrow H^m$, $H^m \rightarrow H^N$, $H^N \rightarrow H^N$, and $H^N \rightarrow H^m$). (*Figure 4.1*)

Recently, Krejcirikov & Tugarinov demonstrated that backbone and methyl resonances can be assigned with a single, uniformly ^{13}C enriched sample[144], and under favorable conditions, the same sample may be employed for collecting distance constraints. We recently presented a so-called time-shared experiment[75], which allowed for simultaneous acquisition of HN-TROSY-NOESY and HC-HSQC-NOESY spectra with minimal sensitivity losses and allowed us to reinvest time to optimize resolution. Critically, NOESY cross-peaks are featured in the detected dimension and benefit from maximal resolution. Such an experiment combined with the strategy presented by Krejcirikov & Tugarinov would allow for a complete structural study using a single sample and a minimal number of experiments. Success of the time-shared experiment, however, ultimately hinges upon generation of a U-$^{2}H^{15}N^{12}C$ sample that is labeled with ^{1}H and ^{13}C at methyl groups of ILV (Me-ILV-$^{1}H^{13}C$, so two expensive samples are needed, regardless of sample stability and solubility. Here, we present a time-shared (TS) HN-TROSY/HC-CT-HMQC-NOESY experiment that can be used with uniform ^{13}C labeling, in which methyl multiple-quantum coherences are encoded with ^{13}C chemical shifts in a constant-time manner while ^{15}N single-quantum coherences evolve in a real-time fashion and with TROSY advantages. The impact of time-shared constraints on sensitivity is minimized, notably through a borrowed evolution period for amide groups and synchronized phase handling.

2. Pulse Sequence

We adapted our TS-(HN-TROSY/HC-HSQC)-NOESY experiment into a TS-(HN-TROSY/HC-CT-HMQC)-NOESY experiment featuring constant-time encoding of carbon frequencies. Since the constant-time period is set to 28 ms (proportional to $1/J_{CC}$), it refocuses splitting under J_{CC}. The nOe distance constraints $H^m \rightarrow H^m$, $H^m \rightarrow H^N$, $H^N \rightarrow H^N$, and $H^N \rightarrow H^m$ can therefore be collected on a sample that is U-$^2H^{15}N^{13}C$ Me-ILV-$^1H^{13}C$, alleviating the requirement of generating a U-$^2H^{15}N^{12}C$ Me-ILV-$^1H^{13}C$ sample. We modeled our CT-HMQC building block after the Griesinger[145] CT-HMQC experiment (see *Figure 4.2*). In the CT-HMQC experiment, the magnetization transfer occurs as follows:

$$\underset{\sigma(a)}{H^M_z} \xrightarrow{J_{HC},\ 180^\circ(H_x,C_x)} \underset{\sigma(b)}{-2H^M_x C^M_z} \xrightarrow{90^\circ(C_x)} \underset{\sigma(c)}{2H^M_x C^M_y} \xrightarrow{180^\circ(H_x,H_x,C_x),\ \omega_C} \underset{\sigma(d)}{-2H^M_x C^M_y \cos(\omega_C t_1)}$$

$$\xrightarrow{90^\circ(C_x)} \underset{\sigma(e)}{-2H^M_x C^M_z \cos(\omega_C t_1)} \xrightarrow{J_{HC},\ 180^\circ(H_x,C_x)} \underset{\sigma(f)}{H^M_y \cos(\omega_C t_1)}$$

Longitudinal magnetization on H^M, σ(a), is converted to proton single quantum coherence anti-phase with respect to C^M, $\sigma(b) = -2H^M_x C^M_z$ (b), and after a 90° ^{13}C pulse, into multiple quantum coherence, $\sigma(b) = 2H^M_x C^M_y$. Evolution of the density operator under the chemical shift of proton is twice refocused with 180° ^1H pulses, whereas evolution under ^{13}C chemical shifts provides $\sigma(d) = -2H^M_x C^M_y \cos(\omega_c t_1)$. Using two 180° pulses on ^1H as opposed to one single 180° pulse midway through the constant time period ensures that $2H^M_x C^M_y$ cannot evolve under J_{HC} with protons not involved in this coherence. Following a 90˚ ^{13}C pulse, a single quantum coherence is generated, $\sigma(e) = -2H^M{}_x C^M{}_z$, which can evolve under J_{HC} to end with $\sigma(f) = H^M{}_y$.

As Ollerenshaw, Tugarinov, and Kay noted, HMQC-like experiments that impose only inversion pulses to methyl groups prevent mixing of slowly and quickly relaxing components of $^{13}C^1H_3$ spin systems, a necessary condition for the so-called “methyl TROSY” effect[146]. In contrast, 90˚ proton pulses and purge gradients used in single quantum coherence experiments result in a density matrix operator reporting on a larger fraction of quickly relaxing components. That is, in this pulse sequence, methyl protons should be subject to an inversion at point *b*. In contrast, to benefit from a TRSOY effect, amide protons must be longitudinal during nitrogen frequency encoding. Ergo, in an effort to create minimal spin manipulation on the HC pathway, so as to take advantage of the slowly relaxing component of the methyl spin system and successfully store amide magnetization longitudinally, methyl and amide coherence pathways were kept orthogonal to one another after the initial period (*Figure 4.3, time points a & b*). Thus, concatenation of a hard 90˚ ^{1}H pulse with a shaped 90˚ $^1H^{methyl}$ pulse between points *b* and *c*, simultaneously created amide longitudinal two-spin order while methyl coherences are only subject to an inversion. Omission of the 90˚ $^1H^{methyl}$ shaped pulse between points *b* and *c* (density matrix operators were of the form $2H^N{}_z N_z$ and $2H^M{}_z C^M{}_z$) and use of a high-powered gradient pulse to destroy any unwanted coherences, resulted in signal losses of 11%, in line with the observations of Ollerenshaw et al.[146] Similarly, the proton shaped pulse between points d_C and e_C and the proton hard pulse between points e_N and *f* cumulate to an inversion for methyl groups while amides are only subject to an excitation. However, here methyl groups are briefly stored as longitudinal two-spin order (see below) so that our experiment does not provide a full methyl-TROSY effect, but it nevertheless minimizes losses. Frequency encoding of heteronuclei requires real-time evolution for ^{15}N and constant-time evolution for ^{13}C. Because the labeling scheme requires refocusing evolution under J_{NCO} and $J_{NC\alpha}$, the easiest solution is to encode these nuclei sequentially, as shown in *Figure 4.3 panel A*. However, methyl groups are subject to unnecessary longitudinal relaxation during t_1, as ^{15}N is encoded between points d_N and e_N. In addition, amide groups suffer from longitudinal relaxation losses during $2\tau_C$ with such a sequential chemical shift encoding. Alternatively, ^{15}N real-time evolution could be constrained within the constant-time period (not shown). Unfortunately, the spectral widths of ^{13}C methyl (25 ppm) and ^{15}N amide groups (35 ppm) impose that dwell times are mismatched by a factor of ~1.8, which would unacceptably limit the resolution in the ^{13}C dimension. Finally, the scheme of *Figure 4.3 panel B* (shown here for t_{1max}) could be employed. Here, methyl groups only relax during a period of (t^{N}_{1max}-$2\tau_C$) and amide relaxation is lessened as ^{15}N frequency is encoded (just as in the previous section). Unfortunately, to prevent first order correction, ^{15}N single quantum coherences would suffer from unnecessary transverse relaxation during 4.55 ms (in contrast to 1.55 ms in other schemes). To palliate these seemingly unavoidable drawbacks, we designed a borrowed evolution: The experiment begins as in *Figure 4.3 panel A* until an evolution time $n^*\Delta t_1 > \tau_{Q5} + \tau_{RS}$, when the sequence shown in *Figure 4.3 panel B* can be employed (see figure caption). This strategy combines the optimal features of each scheme (i) minimal losses due to ^{15}N transverse relaxation at t_1~0 and (ii) combatting longitudinal relaxation as encoding evolves toward t_{1max}.

Sensitivity enhanced TROSY encoding of amide groups and quadrature detection for both amide and methyl pathways are accomplished simultaneously as described for the TS-(HN-TROSY/HC-HSQC)-NOESY experiment[75] but requires synchronized phase handling in the preparation period. Briefly, four transients are recorded ($A1_x$-$A4_x$) with $\phi_4 = -y$ and $\phi_5 = x$, and again with $\phi_4 = x$ and $\phi_5 = y$ ($A1_y$-$A4_y$). Post-acquisition processing provides the real and imaginary components in both t_1 and t_2 evolution periods together with line-selection. Similarly, ^{15}N and ^{13}C pathways are delineated with phase alteration of ω_2 and processing. Tables 4.1 and 4.2 describe density matrix operators for the H^NN and H^MC^M coherence pathways, points e_N/e_C through k. Table 4.3 shows the transients recorded. This so-called nested acquisition is a priori incompatible with the experiment described so far. Indeed, the phase of ϕ_1 alternates between y (for A1 and A3) and x (for A2 and A4), so that the corresponding pulse does not always lead to an inversion when combined with the preceding methyl selective pulse. This impediment leads to 7-14% losses for methyl signals. This drawback can be overcome by adjusting the phases ϕ_6, ϕ_7, and ϕ_8 as described in *Figure 4.3* and in *Tables 4.1 & 4.2*, which simultaneously ensures that methyl are subject to (pseudo) inversions while amide groups benefit from TROSY (including addition of ^{15}N magnetization).

3. Results

We compare our time-shared pulse sequence to the two experiments that it subsumes (HN-TROSY-NOESY and HC-CT-HMQC-NOESY) and also with its predecessor experiment (TS-(HN-TROSY/HC-HSQC)-NOESY). All data were collected on the yersiniabactin synthetase[103–105] U-$^{13}C^{15}N^{2}H$, Me-ILV-^{1}H 52 kDa Cy1 domain[75,138,147] (445 residues). *Figure 4.8* showcases the well-known advantage of constant-time evolutions for uniformly labeled samples and demonstrates integration of the time-shared principle. Here, H/C planes of time-shared experiments are compared for real-time evolution (Mishra et al.[75], purple) and constant-time evolution (pulse sequence in *Figure 1 panels A* & *B*, blue). Clearly, real-time evolution leads to signal-splitting and/or line-broadening that deteriorate the resolution of the spectrum and exacerbate signal overlap. The resolution is rescued when constant-time evolution is employed, yet sensitivity is compromised (as anticipated for constant-time experiments).

Comparison between non time-shared HN-TROSY-NOESY and the corresponding time-shared experiment reveals a reduction in about 12% in signal-to-noise (*Figure 4.6*). The decrease in sensitivity originates from longitudinal relaxation of amide groups during $2\tau_C$; when the constant-time is removed from the experiment in *Figure 4.1 panel A*, which can now be recorded up to t_{1max}, we observed no appreciable losses, in agreement with our previous observations[75]. Similarly, when the constant-time period is included and experiment in *Figure 4.1 panel A* is collected up to t_{1max} (sequential evolution, see above), we observe 5-15% losses in sensitivity. When we overlay our H^{noe}/C^{methyl} and $H^{noe}/H^{indirect}$ planes with those of an HC-CT-HMQC-

NOESY, derived from the experiment of Griesinger et al.[145] we notice ~ 35% loss in signal intensity (*Figure 4.7*). The time-shared INEPT poses a penalty of 21%, as previously observed. The remaining losses are due to the 90˚ (10%) and 180˚ (7%) proton shaped-pulses, and the 90˚ ^{13}C shaped pulse (12%). Secondly, by virtue of the time-shared principle, transfers times for $H^N \rightarrow N$ coherence and $H^M \rightarrow 2H^MC^M$ coherence cannot be optimized for both groups, simultaneously. We have optimized our time-shared INEPT transfers for $H^N \rightarrow N$ transfer; thus results are biased for optimal sensitivity of H^N source peak nOes. Finally, in the time-shared experiment, the methyl interferogram is subject to a relaxation of $e^{(-R_1 t_{III}^N)}$ which is absent in the reference experiment.

Structural studies of larger proteins by NMR spectroscopy rely heavily on selective protonation of ILV methyls in a perdeuterated background. Since these strategies absolutely require generation of a U-$^{13}C^{15}N^2H$ Me-ILV-^1H for methyl resonance assignment, for stable and soluble samples where sensitivity is not an issue, this TS-(HN-TROSY/HC-CT-HMQC)-NOESY experiment may provide distance constraints necessary for structure determination without the need of generating a costly U-$^{12}C^{15}N^2H$, Me-ILV-$^1H^{13}C$ sample. Because such samples have been shown to be amenable for backbone resonance assignment, a single sample may be used for a complete study. In general, we encourage probing the sensitivity of this TS-(HN-TROSY/HC-CT-HMQC)-NOESY experiment following typical methyl assignment pulse sequences

4. Materials and methods

a) Expression & purification of U-$^{2}H^{13}C^{15}N$-MeILV-$^{1}H^{13}C$ Cy1

Experimental data were recorded on a 52 kDa cyclization domain (Cy1) of the HMPW2 subunit of yersiniabactin synthetase[103]. Expression and purification of this domain, and preparation of NMR samples were achieved as previously described[75] with the following exceptions. A subculture of BL21(DE3) *E. coli* cells harboring the pET30a/Cy1 overexpression plasmid were grown overnight in LB supplemented with Kanamycin (50 μg L^{-1}) prepared in H_2O. After ~ 16 hours of overnight expression, 1 mL of overnight culture was used to inoculate 1 L of minimal medium prepared in 99.9% $^{2}H_2O$ solvent: 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 2 mM $MgCl_2$, 0.1 mM CaCl2, 2 g ^{2}H-glucose, 1g ^{15}NH4Cl, 10 mL vitamin solution (prepared in 99.9% $^{2}H_2O$: 0.5 g L^{-1} thiamine, 0.1 g L^{-1} D-biotin, 0.1 g L^{-1} choline chloride, 0.1 g L^{-1} folic acid, 0.1 g L^{-1} niacinamide, 0.1 g L^{-1} D-panthothenic acid, 0.1 g L^{-1} pyridoxal and 0.01 g L^{-1} riboflavin, 2 mL trace element solution[148] (prepared in 99.9% $^{2}H_2O$), and 50 μg L^{-1} of kanamycin). Upon $O.D._{600}$ ~ 0.6, cells were chilled to about 16˚C (ice bath) and shaker temperature was lowered to 16˚C. Metabolic precursors (dissolved in 5 mL $^{2}H_2O$) were added: 62.5 mg ^{1}H-Me-$^{13}C_4$-α-ketobuytyrate-3,3,-^{2}H (CDLM-4611) and 125 mg ^{1}H-Me-$^{13}C_5$-α-ketoisovalerate-1-^{2}H (CDLM-4418), and protein expression was induced with 0.5 mM IPTG for an additional 12 hours, until $O.D._{600}$ reached ~ 1.4. Protein purification and sample preparation was handled as described previously.

b) Acquisition and Processing

All NMR experiments were recorded at 298 K on a Bruker 600 MHz AVANCE III spectrometer equipped with a QCI cryoprobe^{TM}. The TS-HN-TROSY/HC-CT-HMQC-NOESY was recorded with 512 (^{1}H detected, nOe) x 50 (^{1}H indirect) x 84 (^{15}N/^{13}C indirect) complex points. The spectral widths in the ^{1}H detected (nOe) and in the indirect ^{15}N/^{13}C dimensions were 16 ppm and 35/24 ppm, respectively. The ^{13}C and ^{15}N carriers were set to 16.3 ppm and 119 ppm, respectively. The indirect ^{1}H dimension spectral width was 6.5 ppm. 16 repetitions of the experiment were stored individually and each accumulating 4 scans. A recycle delay of 1 s was used, which resulted in a total experimental time of 4 days and 2 hours.

The data are initially processed with a python script (available on request) that handles the necessary signal algebra to separate the interleaved ^{15}N and ^{13}C pathways according to *Table 4.3* and arranges data such that NMRPipe[149] scripts can be used for processing. The detected dimensions of data sets were processed with cosine-squared bell apodization functions and zero filled to 1,024 points, before Fourier Transform. Linear prediction was used in all indirect dimensions to double the number of points, then indirect dimensions were zero-filled to 512 points and 256 points in the ^{15}N/^{13}C and ^{1}H dimensions, respectively, before Fourier Transform was applied.

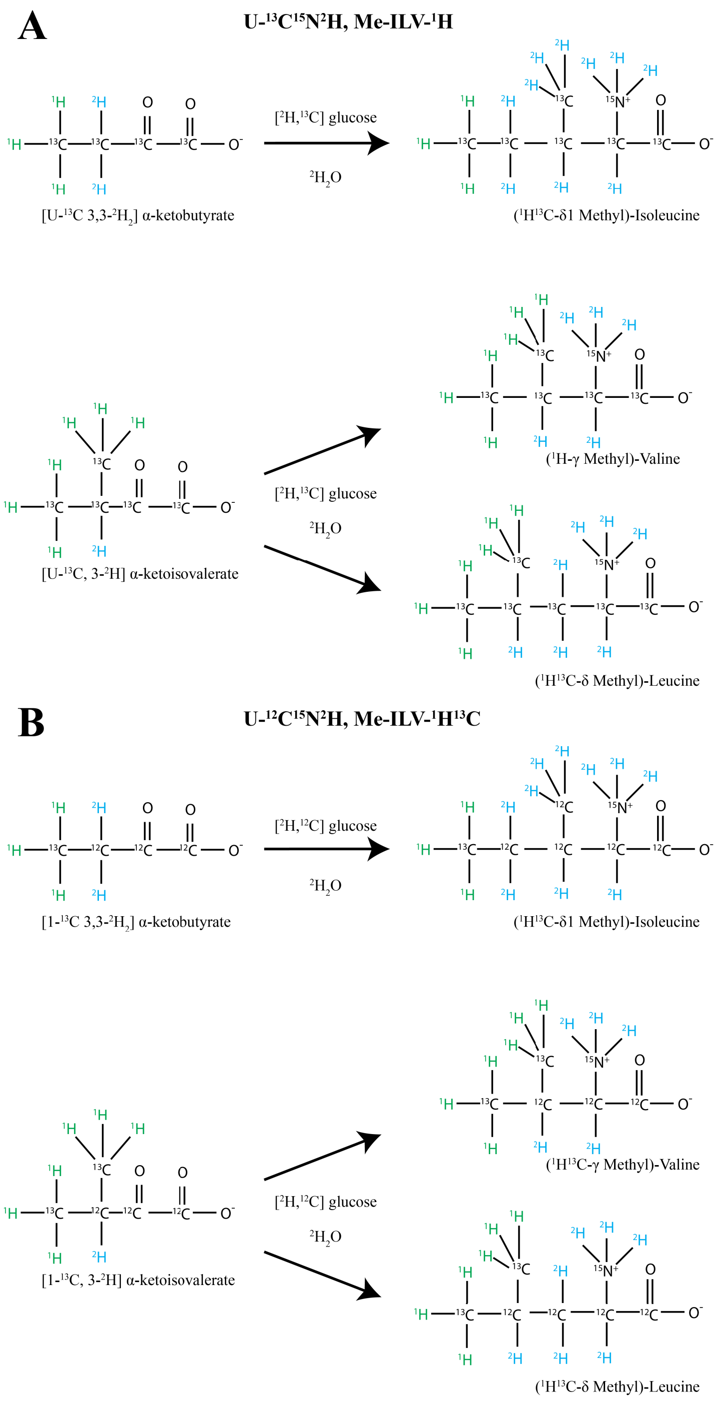
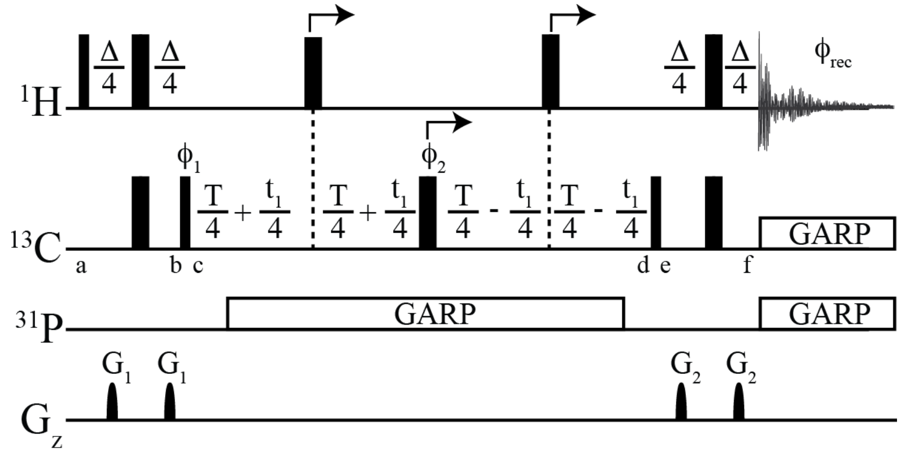


Figure 4.1 Two samples are required for ^{1}H Methyl ILV labeling for resonance assignment. (A) U-^{13}C^{15}N^{2}H, Me-ILV-^{1}H for methyl assignment, and (B) U-^{12}C^{15}N^{2}H, Me-ILV-^{1}H^{13}C for nOe data collection



$$\underset{\sigma(a)}{H^M_z} \xrightarrow{J_{HC},\ 180^\circ(H_x,C_x)} \underset{\sigma(b)}{-2H^M_xC^M_z} \xrightarrow{90^\circ(C_x)} \underset{\sigma(c)}{2H^M_xC^M_y} \xrightarrow{180^\circ(H_x,H_x,C_x),\ \omega_C} \underset{\sigma(d)}{-2H^M_xC^M_y\cos(\omega_C t_1)}$$

$$\xrightarrow{90^\circ(C_x)} \underset{\sigma(e)}{-2H^M_xC^M_z\cos(\omega_C t_1)} \xrightarrow{J_{HC},\ 180^\circ(H_x,C_x)} \underset{\sigma(e)}{H^M_y\cos(\omega_C t_1)}$$

Figure 4.2 Pulse sequence of CT-HMQC and density matrix operators at various time points, a through e. This pulse sequence was reproduced from Marino and Griesinger[145], and it is the standard Bruker "CT-HMQC" sequence used for H/C correlation maps (omitting the ^{31}P CPD). ϕ_1 = *x, -x* ; ϕ_2 = *2(x), 2(-x)* ; ϕ_{rec} = *x, -x*

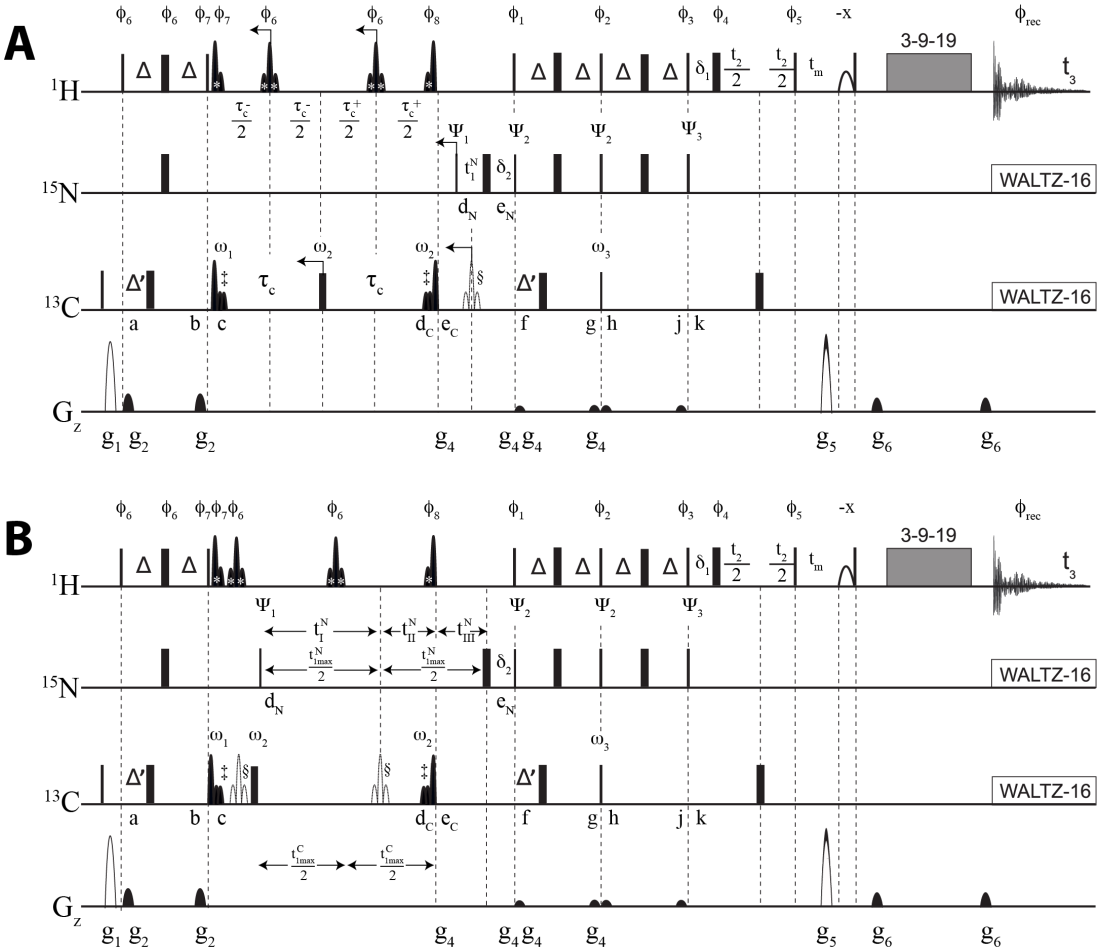
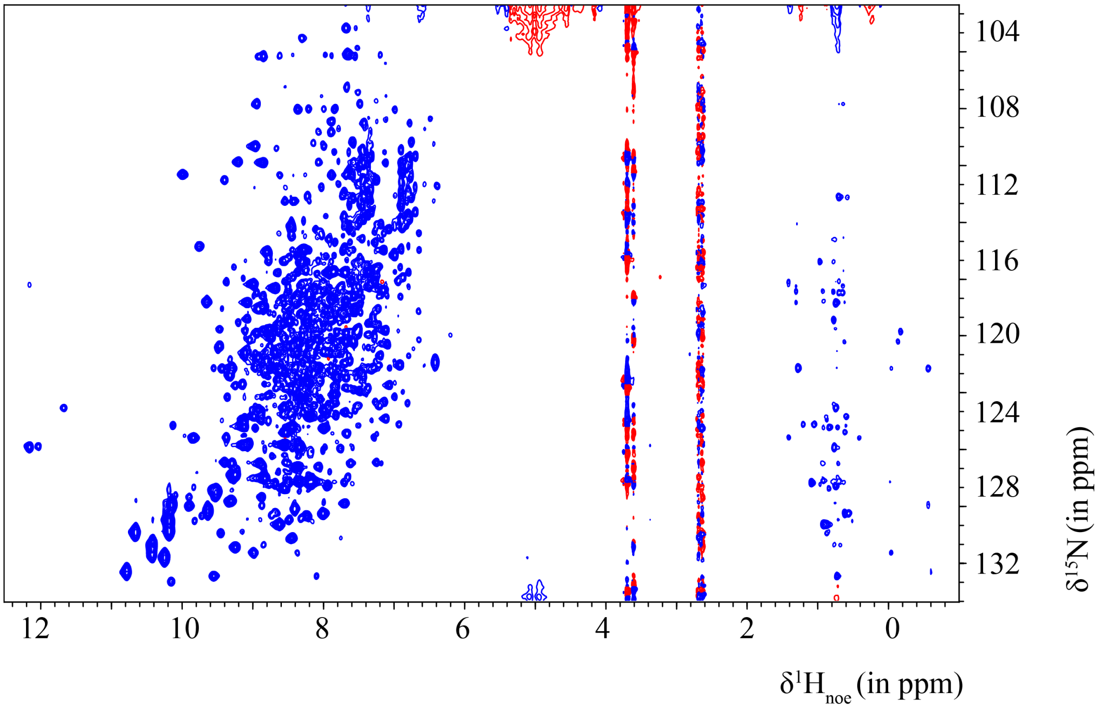


Figure 4.3 TS-(HN-TROSY/HC-CT-HMQC)-NOESY experiment. (**A**) Pulse sequence at $t_1 \sim 0$ (**B**) Pulse sequence at $t_1 \sim 28$ ms (= t_{1max}). The pulse sequence shown in **A** is used until $n^*\Delta t_1 > \tau_{Q5} + \tau_{RS}$ (defined below), at which point, **B** is used so that the Rsnob pulse is accommodated by the constant time period, who now also includes a Bloch-Siegert compensation pulse. Density matrix operators are described for time points *a* through *k* are described in *Tables 4.1* & *4.2*; since the relative positions of some pulses change between **A** and **B**, time points *e* and *f* have subscripts in order to report independently on the HN or HC coherence pathways (*Tables 4.1* & *4.2*). Narrow and wide bars represent 90˚ and 180˚ pulses, respectively. All pulses are applied along the *x*-axis, unless otherwise specified; phases are detailed in

accordance with the spin-dynamic nutation axes detailed in Roehrl et al.[150] The empty ellipsoid represents a water selective 90˚ rectangular pulse (1 ms). The shaped pulses labeled *inside the shape* with a single asterisk (*) are Eburp2 ^{1}H 90˚ methyl selective pulses (τ_{EB} = 1.9 ms), and the shaped pulses labeled *inside the shape* with a double asterisk (**) are Reburp ^{1}H 180˚ methyl selective pulses (τ_{RB} = 2.5 ms). The pulses labeled with a double dagger (‡) are Q5 ^{13}C methyl selective 90˚ pulses (τ_{Q5} = 1.5 ms), and the pulses labeled with § are amplitude-modulated Rsnob pulses (τ_{RS} = 775 μs) that cover both C^{α} and CO frequencies. The block labeled as 3-9-19 is a WATERGATE water suppression scheme[151]. The delays are $\Delta = (1/4J_{NH})$, $\Delta' = (1/4J_{CH})$, and t_m (mixing time) = 150 ms. The delays δ1 (= length of 180˚ ^{13}C rectangular pulse) and δ2 (= τ_{RS}) are used to prevent first-order phase corrections in indirect dimensions. The total constant time period $2\tau_c$ is equal to 28 ms. The delays $\frac{\tau_C^-}{2} = \frac{1}{4\,J_{CC}}$ and are decremented by $\frac{\Delta t_1^C}{4}$ where $\Delta t_1^C = \frac{1}{SW_C}$ is the dwell-time for the ^{13}C dimension; likewise, $\frac{\tau_C^+}{2} = \frac{1}{4\,J_{CC}}$ and are incremented by $\frac{\Delta t_1^C}{4}$. Nitrogen frequencies are encoded with t_1^N and the increment $\Delta t_1^N = \frac{1}{SW_N}$. The pulse sequence shown in **B** is used when $t_1 = n*\Delta t_1 > \tau_{Q5} + \tau_{RS}$; here, t_1 is incremented such that encoding principally takes place during $2\tau_c$ while releasing constraints on the size of SW_N and the resolution of ^{13}C. In both **A** and **B**, the Rsnob pulse (§) is applied halfway through t_1 to refocus evolution under J_{NCO} and $J_{NC\alpha}$. Carbon frequency encoding in **B** follows what is described for **A**. Practically, encoding of ^{15}N in **B** is achieved with three separate delays: t_I^N, t_{II}^N, and t_{III}^N with increments $\Delta t_I^N = \frac{1}{2SW_N}$, $\Delta t_{II}^N = \frac{\tau_C}{2N_{max}}$, and $\Delta t_{III}^N = \frac{1}{2SW_N}$. Most of the encoding of ^{15}N, occurs during the ^{13}C constant-time period ($2\tau_c$), such that at t_{1max}^{N}, $t_{III}^{N} = t_{1max}^{N} - 2\tau_{C}$. At The filled ellipsoids on the line labeled G_Z are 1 ms smoothed-square shaped-pulse gradients: g_1 = -36 G/cm, g_2 = 7.5 G/cm, g_4 = 2.5 G/cm, g_5 = -40 G/cm, g_6 = 4 G/cm; each gradient pulse is followed by a 200 µs recovery delay. ^{13}C and ^{15}N decoupling during acquisition are both achieved using WALTZ-16 sequences[152], with field strengths of 0.71 kHz. Simultaneous ^{13}C and ^{15}N decoupling necessitates the use of lower field strengths and synchronous pulsing to prevent artifacts[153]. Quadrature detection in both indirect dimensions (t_1 and t_2) is achieved by post-acquisition combinations of transients that are recorded and stored separately, which have the following phase settings: **A_1**: $\phi_1 = y$, $\phi_2 = x$, $\phi_3 = -x$, $\phi_6 = x$, $\phi_7 = -y$, $\phi_8 = y$, $\psi_1 = y\ {-y}$, $\psi_2 = -y$, $\psi_3 = x$, $\omega_1 = -x\ x$, $\omega_2 = x$, $\omega_3 = x\ x\ {-x}\ {-x}$; **$A_2$**: $\phi_1 = x$, $\phi_2 = y$, $\phi_3 = -y$, $\phi_6 = y$, $\phi_7 = x$, $\phi_8 = x$, $\psi_1 = y\ {-y}$, $\psi_2 = -x$, $\psi_3 = y$, $\omega_1 = -y\ y$, $\omega_2 = x$, $\omega_3 = x\ x\ {-x}\ {-x}$; **A_3**: $\phi_1 = y$, $\phi_2 = x$, $\phi_3 = x$, $\phi_6 = x$, $\phi_7 = -y$, $\phi_8 = y$, $\psi_1 = y\ {-y}$, $\psi_2 = -y$, $\psi_3 = x$, $\omega_1 = -x\ x$, $\omega_2 = x$, $\omega_3 = x\ x\ {-x}\ {-x}$; **A_4**: $\phi_1 = x$, $\phi_2 = y$, $\phi_3 = y$, $\phi_6 = y$, $\phi_7 = x$, $\phi_8 = x$, $\psi_1 = y\ {-y}$, $\psi_2 = -x$, $\psi_3 = y$, $\omega_1 = -y\ y$, $\omega_2 = x$, $\omega_3 = x\ x\ {-x}\ {-x}$. $\phi_{rec} = x, -x$ for all transients. A1-A4 are recorded with $\phi_4 = -y$ and $\phi_5 = x$ ($A1_x$-$A4_x$) or $\phi_4 = x$ and $\phi_5 = y$ ($A1_y$-$A4_y$), where the subscript denotes the phase of ϕ_5. To delineate ^{15}N and ^{13}C edited spectra, these 8 transients were recorded a second time with the phase ω_2 inverted. TPPI[154] is implemented during t1 evolution using the phases ψ_1, ω_1, and ϕ_{rec}. All experiments were conducted at 298K on a 600 MHz Bruker Avance III spectrometer equipped with a CP-QCI ^{1}H/^{31}P/^{13}C/^{15}N-^{2}H cyroprobe with a single axis gradient coil. The Bruker pulse program can be obtained by contacting the corresponding author.


HN pathway: $^{1}H_{noe}/^{15}N$ (^{13}C) projection

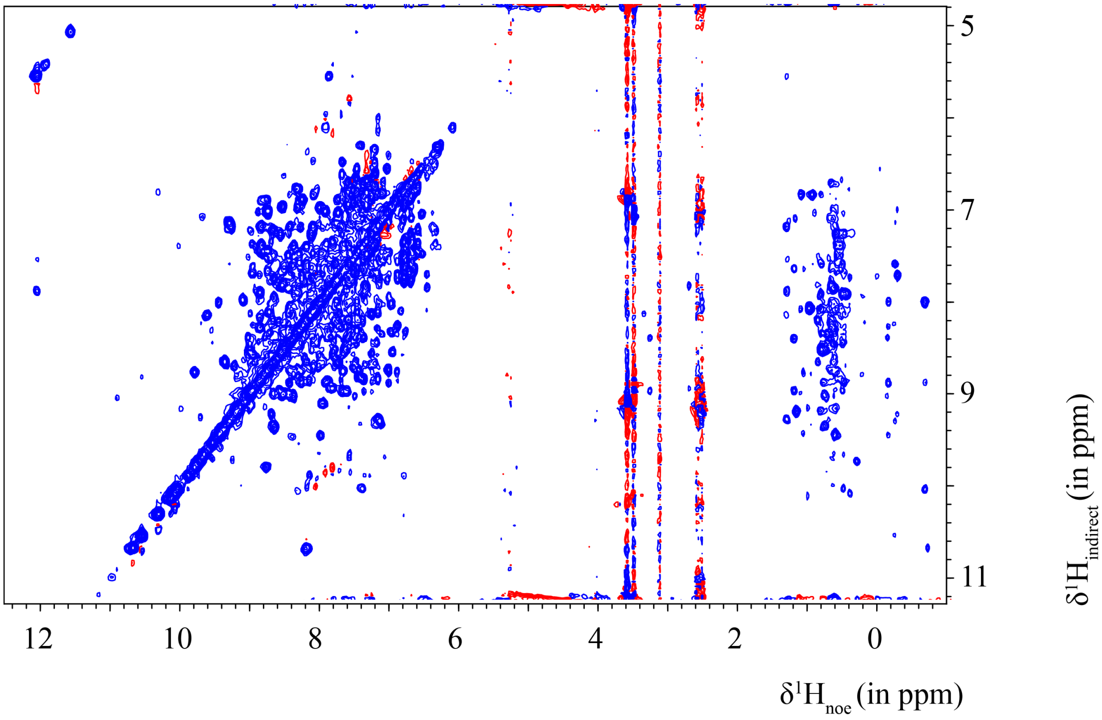
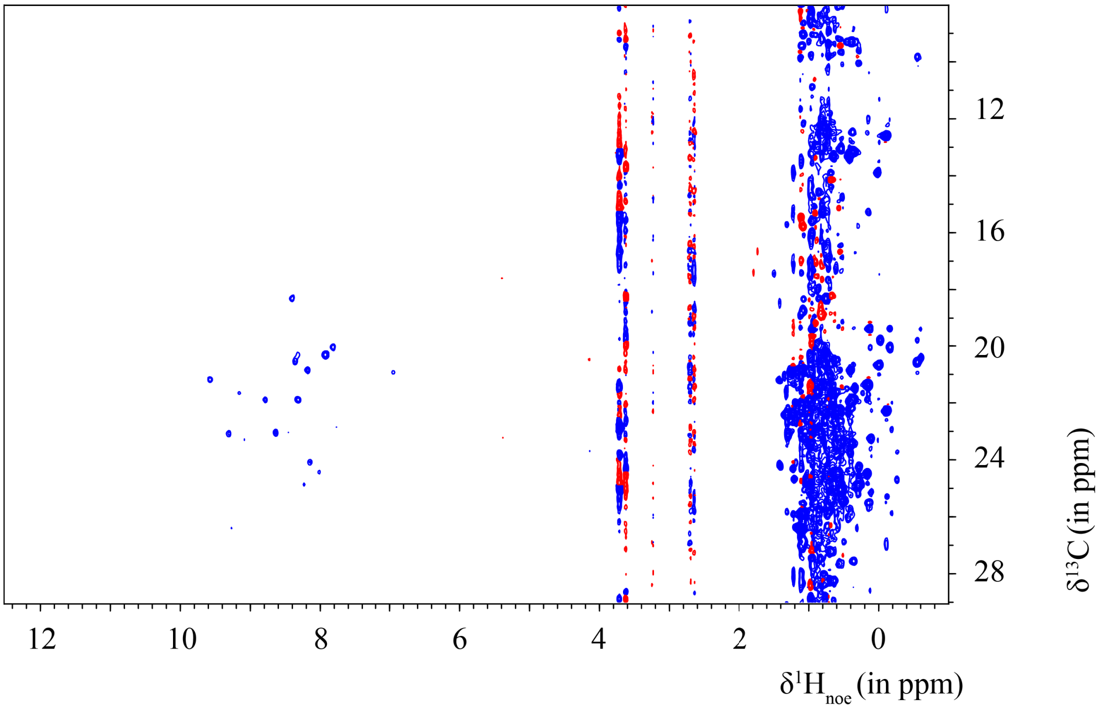

HN pathway: $^{1}H_{noe}/^{1}H_{indirect}$ projection

Figure 4.4 2D projections of 3D TS-(HN-TROSY/HC-CT-HMQC)-NOESY spectrum obtained with the pulse sequence shown in Figure 4.3, HN pathway

subspectra. Cy1 was prepared in 20 mM Sodium Phosphate, 10 mM NaCl, 1 mM EDTA, 5 mM DTT, 0.05% Sodium Azide (w/v), pH 7.0, 5% ^2H_2O (for lock). Data were collected at 298 K on a 600 MHz spectrometer. 16 repetitions of the experiments were stored individually, with each accumulating 4 scans (using a one second recycling delay). The mixing time was 150 ms. 512 complex points were acquired in the direct ^1H dimension, 50 complex points were acquired in the indirect ^1H dimension, and 84 complex points were acquired in the shared $^{15}N/^{13}C$ dimension.

HC pathway: ${}^{1}H_{noe}/{}^{13}C$ (${}^{15}N$) projection



HC pathway: ${}^{1}H_{noe}/{}^{1}H_{indirect}$ projection

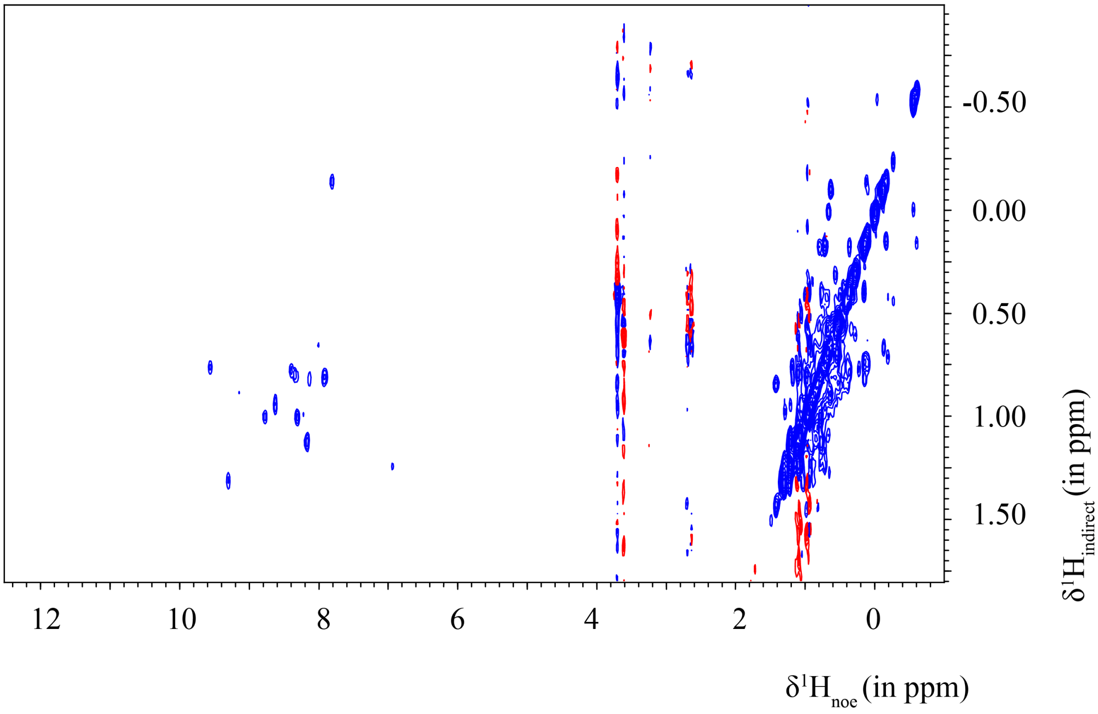


Figure 4.5 2D projections of 3D TS-(HN-TROSY/HC-CT-HMQC)-NOESY spectrum obtained with the pulse sequence shown in Figure 4.3, HC pathway

subspectra. Cy1 was prepared in 20 mM Sodium Phosphate, 10 mM NaCl, 1 mM EDTA, 5 mM DTT, 0.05% Sodium Azide (w/v), pH 7.0, 5% ^2H_2O (for lock). Data were collected at 298 K on a 600 MHz spectrometer. 16 repetitions of the experiments were stored individually, with each accumulating 4 scans (using a one second recycling delay). The mixing time was 150 ms. 512 complex points were acquired in the direct ^1H dimension, 50 complex points were acquired in the indirect ^1H dimension, and 84 complex points were acquired in the shared $^{15}N/^{13}C$ dimension.

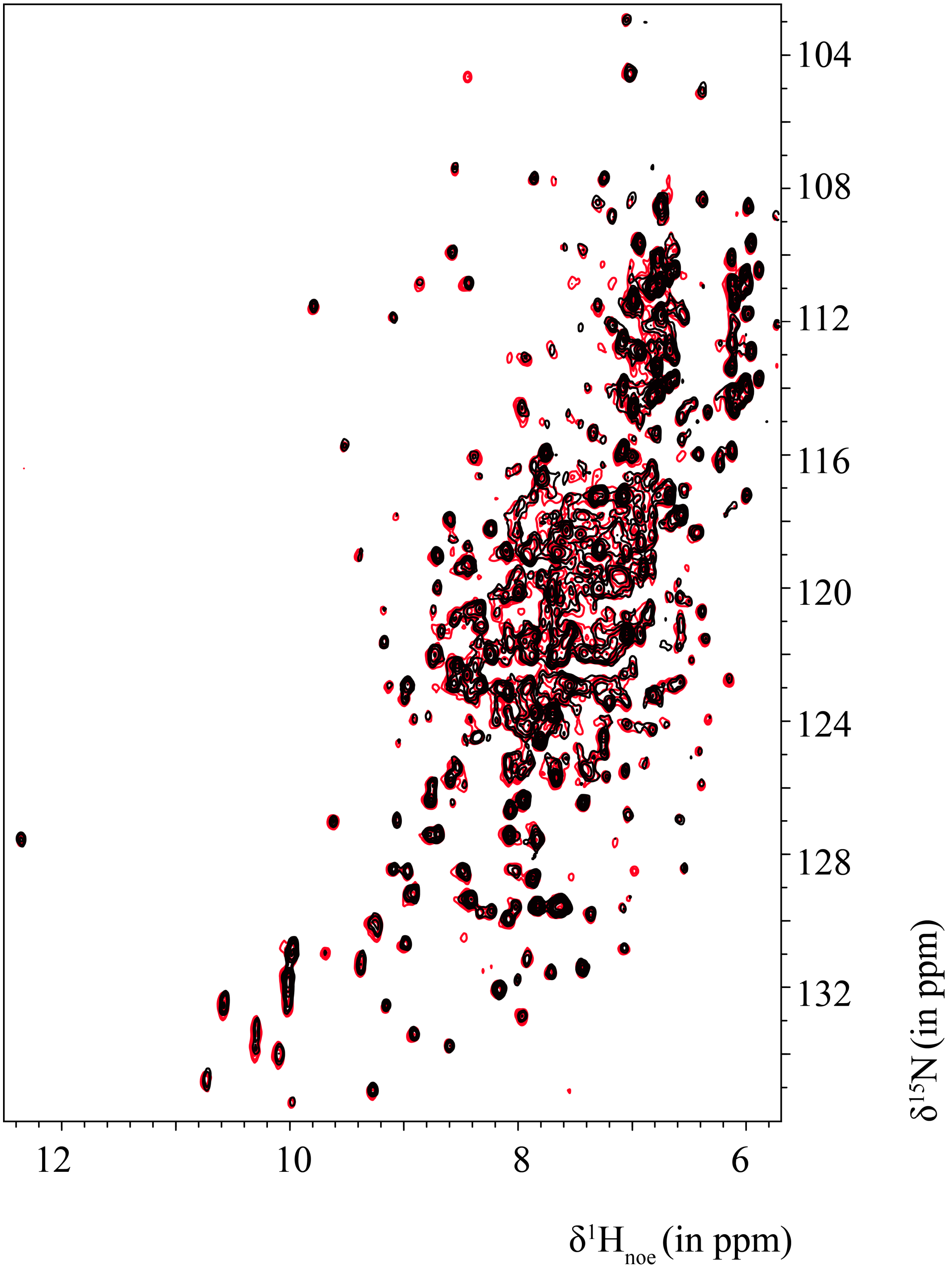


Figure 4.6 TS-(HN-TROSY/HC-HSMC)-NOESY comparison with HN-TROSY-NOESY. The time-shared experiment is shown in **black** and the non-time shared version is shown in **red**. Cy1 was prepared in 20 mM Sodium Phosphate, 10 mM NaCl, 1 mM EDTA, 5 mM DTT, 0.05% Sodium Azide (w/v), pH 7.0, 5% ^2H_2O (for

lock). Data were collected at 298 K on a 600 MHz spectrometer. 16 repetitions of the experiments were stored individually, with each accumulating 4 scans (using a one second recycling delay). The mixing time was 150 ms. 512 complex points were acquired in the direct ^{1}H dimension, 50 complex points were acquired in the indirect ^{1}H dimension, and 84 complex points were acquired in the shared $^{15}N/^{13}C$ dimension.

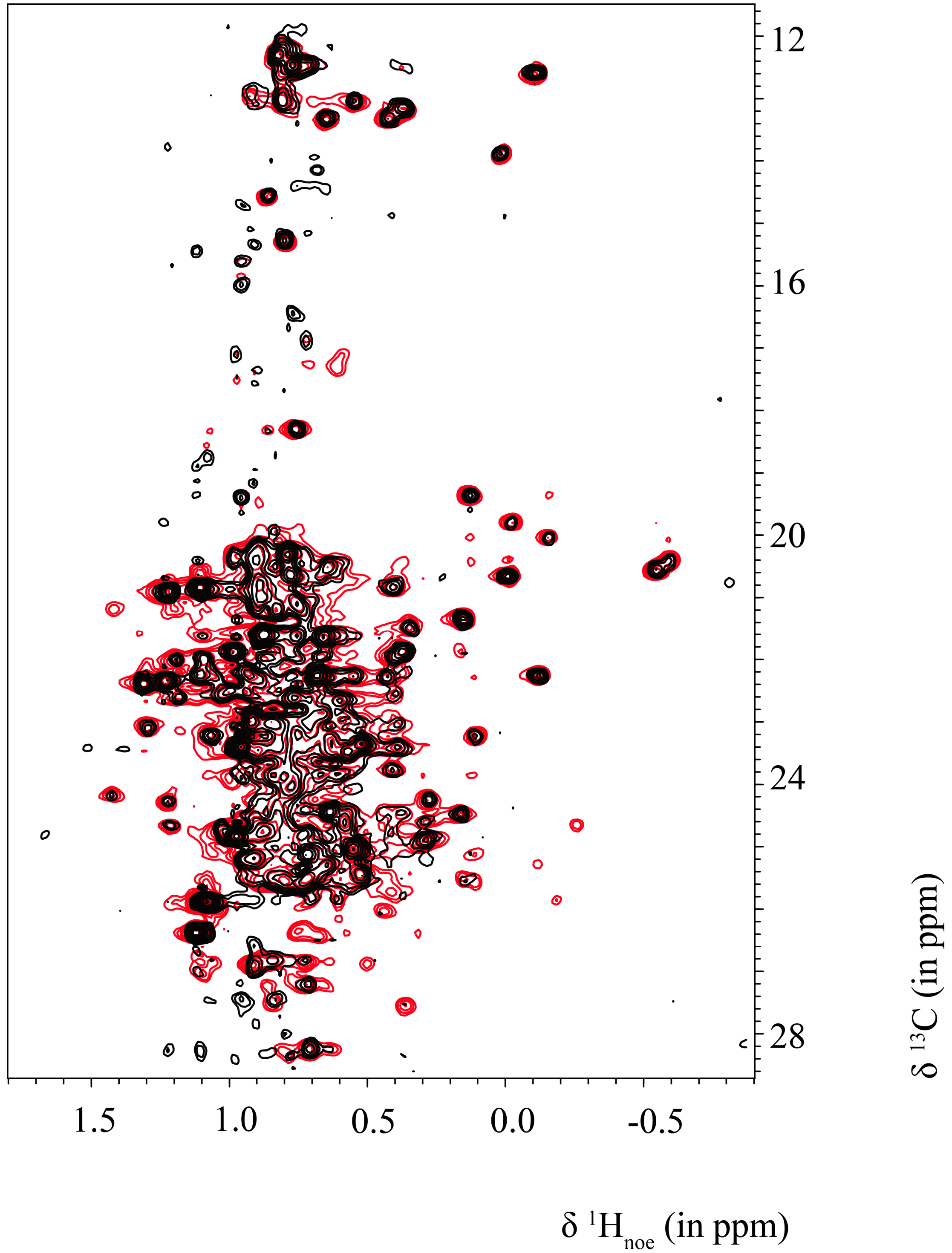


Figure 4.7 TS-(HN-TROSY/HC-CT-HMQC)-NOESY comparison with CT-HMQC-NOESY. The time-shared experiment is shown in **black** and the non-time shared version is shown in **red**. Cy1 was prepared in 20 mM Sodium Phosphate, 10

mM NaCl, 1 mM EDTA, 5 mM DTT, 0.05% Sodium Azide (w/v), pH 7.0, 5% ^2H_2O (for lock). Data were collected at 298 K on a 600 MHz spectrometer. 16 scans were collected using a one second recycling delay. The mixing time was 150 ms. 512 complex points were acquired in the direct ^1H dimension and 84 complex points were acquired in the shared ^{15}Ndimension.

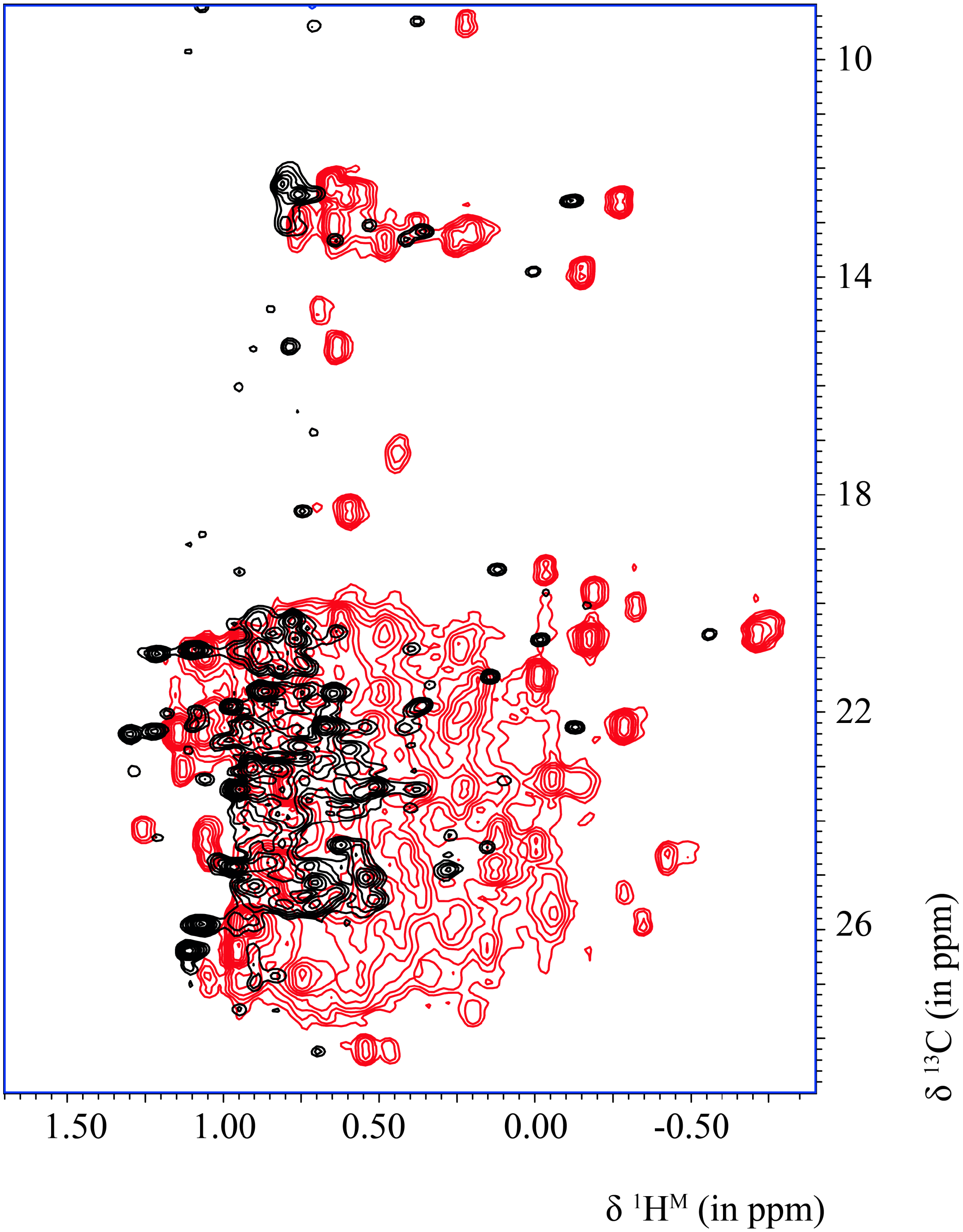


Figure 4.8 TS-(HN-TROSY/HC-HSQC)-NOESY comparison with TS-(HN-TROSY/HC-CT-HMQC)-NOESY. Overlay of HC component of the real time TS-(HN-TROSY/HC-HSQC)-NOESY experiment (**red**) with the HC component of the constant time TS-(HN-TROSY/HC-CT-HMQC)-NOESY experiment (**black**). The ^{1}H^{M} dimension of TS-(HN-TROSY/HC-HSQC)-NOESY spectrum (**red**) has been

shifted upfield by 0.25 ppm to emphasize splitting of lines under J_{CC}. In each experiment, the same number of indirect dimension points were collected in the ^{13}C dimension (84 points, which corresponds to ~ 28 ms). Note that the real time HC-HSQC experiment exhibits line splitting due to J_{CC}, which does not occur in the HC-CT-HMQC experiment. Cy1 was prepared in 20 mM Sodium Phosphate, 10 mM NaCl, 1 mM EDTA, 5 mM DTT, 0.05% Sodium Azide (w/v), pH 7.0, 5% $^{2}H_2O$ (for lock). Data were collected at 298 K on a 600 MHz spectrometer. 16 scans were collected, using a one second recycling delay. The mixing time was 150 ms. 512 complex points were acquired in the direct ^{1}H dimension, and 84 complex points were acquired in the shared ^{13}C dimension.

	a	**b**	**c**	**d$_N$**	**e$_N$**	**f**	**g**	**h**	**j**	**k**	
A1	$-H^N_y$	$2H^N_xN_z$	$2H^N_zN_z$	$-2H^N_zN_x$	$2H^N_zN_x$	$-2H^N_xN_z$	$-H^N_y$	$-H^N_z$	H^N_z	H^N_y	½ $({}^{N}c^{+} + {}^{N}c^{-})$
	N_z	$-N_z$	$-N_z$	N_x	N_x	$-N_z$	N_z	N_x	$-2H^N_zN_y$	$2H^N_yN_z$	½ $({}^{N}c^{-} - {}^{N}c^{+})$
					$2H^N_zN_y$	$2H^N_xN_y$	$-2H^N_xN_y$	$-2H^N_xN_y$	$2H^N_xN_y$	$-2H^N_xN_z$	½ $({}^{N}s^{+} + {}^{N}s^{-})$
					N_y	N_y	$-2H^N_zN_x$	$-2H^N_yN_z$	$-H^N_x$	$-H^N_x$	½ $({}^{N}s^{+} - {}^{N}s^{-})$
A2	$-H^N_x$	$2H^N_yN_z$	$2H^N_zN_z$	$-2H^N_zN_x$	$2H^N_zN_x$	$-2H^N_yN_x$	$2H^N_yN_x$	$2H^N_yN_x$	$-2H^N_yN_x$	$-2H^N_yN_z$	½ $({}^{N}c^{+} + {}^{N}c^{-})$
	N_z	$-N_z$	$-N_z$	N_x	N_x	N_x	$-2H^N_zN_y$	$-2H^N_xN_z$	$-H^N_y$	$-H^N_y$	½ $({}^{N}c^{-} - {}^{N}c^{+})$
					$2H^N_zN_y$	$-2H^N_yN_z$	$-H^N_x$	$-H^N_z$	$-H^N_z$	H^N_x	½ $({}^{N}s^{+} + {}^{N}s^{-})$
					N_y	N_z	$-N^-$	$-N_y$	$-2H^N_zN_x$	$2H^N_xN_z$	½ $({}^{N}s^{+} - {}^{N}s^{-})$
A3	$-H^N_y$	$2H^N_xN_z$	$2H^N_zN_z$	$-2H^N_zN_x$	$2H^N_zN_x$	$-2H^N_xN_z$	$-H^N_y$	$-H^N_z$	H^N_z	$-H^N_y$	½ $({}^{N}c^{+} + {}^{N}c^{-})$
	N_z	$-N_z$	$-N_z$	N_x	N_x	$-N_z$	N_z	N_x	$-2H^N_zN_y$	$-2H^N_yN_z$	½ $({}^{N}c^{-} - {}^{N}c^{+})$
					$2H^N_zN_y$	$2H^N_xN_y$	$-2H^N_xN_y$	$-2H^N_xN_y$	$2H^N_xN_y$	$-2H^N_xN_z$	½ $({}^{N}s^{+} + {}^{N}s^{-})$
					N_y	N_y	$-2H^N_zN_x$	$-2H^N_yN_z$	$-H^N_x$	$-H^N_x$	½ $({}^{N}s^{+} - {}^{N}s^{-})$
A4	$-H^N_x$	$2H^N_yN_z$	$2H^N_zN_z$	$-2H^N_zN_x$	$2H^N_zN_x$	$-2H^N_yN_x$	$2H^N_yN_x$	$2H^N_yN_x$	$-2H^N_yN_x$	$-2H^N_yN_z$	½ $({}^{N}c^{+} + {}^{N}c^{-})$
	N_z	$-N_z$	$-N_z$	N_x	N_x	N_x	$-2H^N_zN_y$	$-2H^N_xN_z$	$-H^N_y$	$-H^N_y$	½ $({}^{N}c^{-} - {}^{N}c^{+})$
					$2H^N_zN_y$	$-2H^N_yN_z$	$-H^N_x$	$-H^N_z$	$-H^N_z$	${}^{-}H^N_x$	½ $({}^{N}s^{+} + {}^{N}s^{-})$
					N_y	N_z	$-N_z$	$-N_y$	$-2H^N_zN_x$	$2H^N_xN_z$	½ $({}^{N}s^{+} - {}^{N}s^{-})$

Table 4.1 Density matrix operators of amide (HN-TROSY) coherences at points *a* through *k* in Figure 4.3

This table has been adapted from Mishra et al.[75] J_{HN} refers to the coupling constant between amide proton and nitrogen. The negative gryomagentic ratio of ^{15}N and the negative sign of J_{HN} were taken into account during all calculations. ω_N is the amide nitrogen frequency encoded during t_1.

${}^{N}c^{+} = \cos[(\omega_N + \pi J_{HN})t_1]$, ${}^{N}c^{-} = \cos[(\omega_N - \pi J_{HN})t_1]$,

${}^{N}s^{+} = \sin[(\omega_N + \pi J_{HN})t_1]$, and ${}^{N}s^{-} = \sin[(\omega_N - \pi J_{HN})t_1]$

	a	**b**	**c**	**d$_C$**	**e$_C$**	**f**	**g**	**h**	**j**	**k**	
A1	$-H^M{}_y$	$-2H^M{}_xC^M{}_z$	$2H^M{}_xC^M{}_y$	$2H^M{}_xC^M{}_y$	$-2H^M{}_zC^M{}_z$	$-2H^M{}_xC^M{}_z$	$H^M{}_y$	$H^M{}_z$	$-H^M{}_z$	$-H^M{}_y$	$\cos(\omega_c t_1)$
A2	$H^M{}_x$	$-2H^M{}_yC^M{}_z$	$-2H^M{}_yC^M{}_x$	$-2H^M{}_yC^M{}_y$	$-2H^M{}_zC^M{}_z$	$2H^M{}_yC^M{}_z$	$-H^M{}_x$	$H^M{}_z$	$-H^M{}_z$	$H^M{}_x$	$\sin(\omega_c t_1)$
A3	$-H^M{}_y$	$-2H^M{}_xC^M{}_z$	$2H^M{}_xC^M{}_y$	$2H^M{}_xC^M{}_y$	$-2H^M{}_zC^M{}_z$	$-2H^M{}_xC^M{}_z$	$H^M{}_y$	$H^M{}_z$	$-H^M{}_z$	$H^M{}_y$	$\cos(\omega_c t_1)$
A4	$H^M{}_x$	$-2H^M{}_yC^M{}_z$	$-2H^M{}_yC^M{}_x$	$-2H^M{}_yC^M{}_y$	$-2H^M{}_zC^M{}_z$	$2H^M{}_yC^M{}_z$	$-H^M{}_x$	$H^M{}_z$	$-H^M{}_z$	$-H^M{}_x$	$\sin(\omega_c t_1)$

Table 4.2 Density matrix operators of methyl (HC-CT-HMQC) coherences at points *a* through *k* in Figure 4.3

This table has been adapted from Mishra et al.[75]

J_{HC} refers to the coupling constant between the methyl proton and methyl carbon. ω_C is the methyl carbon frequency encoded during t_1.

	Transient recorded
A1$_x$	$S_N e^{(-i\omega_H t_3)}\left(c^+_N c^-_{HN} - s^+_N s^-_{HN} + c^-_N c^+_{HN} + s^-_N s^+_{HN}\right) - S_C e^{(-i\omega_H t_3)}\cos(\omega_c t_1)\cos(\omega_{HC} t_2)$
A2$_x$	$S_N e^{(-i\omega_H t_3)}\left(-c^+_N c^-_{HN} + s^+_N s^-_{HN} + c^-_N c^+_{HN} + s^-_N s^+_{HN}\right) - S_C e^{(-i\omega_H t_3)}\sin(\omega_c t_1)\sin(\omega_{HC} t_2)$
A3$_x$	$S_N e^{(-i\omega_H t_3)}\left(-c^+_N c^-_{HN} - s^+_N s^-_{HN} - c^-_N c^+_{HN} + s^-_N s^+_{HN}\right) + S_C e^{(-i\omega_H t_3)}\cos(\omega_c t_1)\cos(\omega_{HC} t_2)$
A4$_x$	$S_N e^{(-i\omega_H t_3)}\left(-c^+_N c^-_{HN} - s^+_N s^-_{HN} + c^-_N c^+_{HN} - s^-_N s^+_{HN}\right) + S_C e^{(-i\omega_H t_3)}\sin(\omega_c t_1)\sin(\omega_{HC} t_2)$
A1$_y$	$S_N e^{(-i\omega_H t_3)}\left(c^+_N s^-_{HN} + s^+_N c^-_{HN} + c^-_N s^+_{HN} - s^-_N c^+_{HN}\right) - S_C e^{(-i\omega_H t_3)}\cos(\omega_c t_1)\sin(\omega_{HC} t_2)$
A2$_y$	$S_N e^{(-i\omega_H t_3)}\left(-c^+_N s^-_{HN} - s^+_N c^-_{HN} + c^-_N s^+_{HN} - s^-_N c^+_{HN}\right) + S_C e^{(-i\omega_H t_3)}\sin(\omega_c t_1)\cos(\omega_{HC} t_2)$
A3$_y$	$S_N e^{(-i\omega_H t_3)}\left(-c^+_N s^-_{HN} + s^+_N c^-_{HN} - c^-_N s^+_{HN} - s^-_N c^+_{HN}\right) + S_C e^{(-i\omega_H t_3)}\cos(\omega_c t_1)\sin(\omega_{HC} t_2)$
A4$_y$	$S_N e^{(-i\omega_H t_3)}\left(-c^+_N s^-_{HN} + s^+_N c^-_{HN} + c^-_N s^+_{HN} + s^-_N c^+_{HN}\right) - S_C e^{(-i\omega_H t_3)}\sin(\omega_c t_1)\cos(\omega_{HC} t_2)$

Table 4.3 Transients recorded before separation of amide and methyl signals. *This table has been reproduced from Mishra et al.*[75] and explains the transients recorded during the pulse sequence shown in *Figure 4.2*. (A1$_x$-A4$_x$) with ϕ_4 = -*y* and ϕ_5 = *x*, and four transients are recorded with ϕ_4 = *x* and ϕ_5 = *y* (A1$_y$-A4$_y$). S_N and S_C are the amplitudes of the detected amide and methyl proton signals, respectively, and ω_H is the frequency of the NOESY cross-peak encoded in the detected proton dimension (t_3). $^Nc^+$, $^Nc^-$, $^Ns^+$, and $^Ns^-$ are described in *Table 4.1*. The negative gryomagentic ratio of ^{15}N and the negative sign of J_{HN} were taken into account during all calculations. ω_{HN} and ω_{HC} are the amide proton and methyl proton frequencies encoded in t_2.

$^{HN}c^+ = \cos[(\omega_{HN} + \pi J_{HN})t_2]$, $^{HN}c^- = \cos[(\omega_{HN} - \pi J_{HN})t_2]$,

Bibliography

1. Kelley LA, Mezulis S, Yates C, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat Protoc*. 2015;10(6):845-858. doi:10.1038/nprot.2015-053.
2. Griesinger C, Sattler M. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc*. 1999;34:93-158.
3. Göbl C, Madl T, Simon B, Sattler M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc*. 2014;80C:26-63. doi:10.1016/j.pnmrs.2014.05.003.
4. Frueh DP. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog Nucl Magn Reson Spectrosc*. December 2013. doi:10.1016/j.pnmrs.2013.12.001.
5. Ohki S, Kainosho M. Stable isotope labeling methods for protein NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc*. 2008;53(4):208-226. doi:10.1016/j.pnmrs.2008.01.003.
6. L MJ, Putter I, Jardetzky O. High-Resolution Nuclear Magnetic Resonance Spectra of Selectively Deuterated Staphylococcal Nuclease. *Science (80-)*. 1968;161(3847):1249-1251.
7. Pervushin K V, Wider G, Wüthrich K. Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J Biomol NMR*. 1998;12(2):345-348. doi:10.1023/A:1008268930690.
8. Zhu G. TROSY-based NMR experiments for NMR studies of large biomolecules. *Prog Nucl Magn Reson*. 2008;52:49-68. doi:10.1016/j.pnmrs.2007.10.001.
9. Mobli M, Stern AS, Hoch JC. Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy. *J Magn Reson*. 2006;182(1):96-105. doi:10.1016/j.jmr.2006.06.007.
10. Hyberts SG, Arthanari H, Wagner G. Applications of non-uniform sampling and processing. *Top Curr Chem*. 2012;316:125-148. doi:10.1007/128_2011_187.
11. Orekhov VY, Jaravine V a. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog Nucl Magn Reson Spectrosc*. 2011;59(3):271-292. doi:10.1016/j.pnmrs.2011.02.002.
12. Mobli M, Maciejewski MW, Schuyler AD, Stern AS, Hoch JC. Sparse sampling methods in multidimensional NMR. *Phys Chem Chem Phys*. 2012;14(31):10835-10843. doi:10.1039/c2cp40174f.
13. Barna JCJ, Laue ED, Mayger MR, P WSJ. Exponential Sampling , an Alternative Method for Sampling in Two-Dimensional NMR Experiments. *J Magn Reson*. 1987;77:69-77.
14. Holland DJ, Bostock MJ, Gladden LF, Nietlispach D. Fast multidimensional NMR spectroscopy using compressed sensing. *Angew Chem Int Ed Engl*. 2011;50(29):6548-6551. doi:10.1002/anie.201100440.

15. Kazimierczuk K, Orekhov VY. Accelerated NMR spectroscopy by using compressed sensing. *Angew Chemie Int Ed*. 2011;50(24):5556-5559. doi:10.1002/anie.201100370.
16. Hyberts SG, Robson S a, Wagner G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J Biomol NMR*. 2013;55(2):167-178. doi:10.1007/s10858-012-9698-2.
17. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A*. 1997;94(23):12366-12371. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24947&tool=pmcentrez&rendertype=abstract.
18. Pervushin K. Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys*. 2000;33(2):161-197. http://www.ncbi.nlm.nih.gov/pubmed/11131563.
19. Czisch M, Boelens R. Sensitivity enhancement in the TROSY experiment. *J Magn Reson*. 1998;134(1):158-160. doi:10.1006/jmre.1998.1483.
20. Weigelt J. Single Scan, Sensitivity- and Gradient-Enhanced TROSY for Multidimensional NMR Experiments *J. Am. Chem. Soc.* **1998** , *120* , 10778−10779. *J Am Chem Soc*. 1998;120(48):12706-12706. doi:10.1021/ja9855287.
21. Loria JP, Rance M, Palmer AG. Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J Magn Reson*. 1999;141(1):180-184. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10527755.
22. Zhu G, Kong XM, Sze KH. Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *J Biomol NMR*. 1999;13:77-81. doi:10.1023/A:1008398227519.
23. Yang D, Kay LE. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *J Am Chem Soc*. 1999;121(11):2571-2575. doi:10.1021/ja984056t.
24. Yang DW, Kay LE. Improved 1HN-detected triple resonance TROSY-based experiments. *J Biomol NMR*. 1999;13(1):3-10. doi:10.1023/A:1008329230975.
25. Nietlispach D. Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *J Biomol NMR*. 2005;31(2):161-166. doi:10.1007/s10858-004-8195-7.
26. Meissner A, Duus JØ, Sørensen OW. Spin-State-Selective Excitation. Application for E.COSY-Type Measurement of JHH Coupling Constants. *J Magn Reson*. 1997;97(128):92-97.
27. Sørensen MD, Meissner A, Sørensen OW. Spin-state-selective coherence transfer via intermediate states of two-spin coherence in IS spin systems: Application to E.COSY-type measurement of J coupling constants. *J Biomol NMR*. 1997;10(2):181-186. doi:10.1023/a:1018323913680.
28. Cavanagh J, Fairbrother WJ, Palmer AGI, Rance M, Skelton NJ. *Principles*

and Practice: Protein NMR Spectroscopy. Vol Second Edi. New York, New York: Academic Press; 2007.

29. Crespi HL, Rosenberg RM, Katz JJ. Proton magnetic resonance of proteins fully deuterated except for 1H-leucine side chains. *Science*. 1968;161(843):795-796. doi:10.1126/science.161.3843.795.
30. Crespi HL, Katz JJ. High Resolution Proton Magnetic Resonance Studies of Fully Deuterated and Isotope Hybrid Proteins. *Nature*. 1969;224:560-562. doi:10.1038/224488a0.
31. Grzesiek S, Anglister J, Ren JH, Bax A. 13C Line Narrowing by ZH Decoupling in 2H/13C/1SN- Enriched Proteins. Application to Triple Resonance 4D. *J Am Chem Soc*. 1993;11:4369-4370.
32. Lemaster DM, Richards FM. NMR Sequential Assignment of Escherichia coli Thioredoxin Utilizing Random Fractional Deuteration. *Biochemistry*. 1988;27:142-150.
33. Muhandiram DR, Yamazaki T, Sykes BD, Kay LE. Measurement of 2H T1 and Tlp Relaxation Times in Uniformly 13C-Labeled and Fractionally 2H-Labeled Proteins in Solution. *J Am Chem Soc*. 1995;117:11536-11544.
34. Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *J Am Chem Soc*. 1994;116:11655-11666. doi:10.1021/ja00105a005.
35. Yamazaki T, Lee W, Revington M, et al. An HNCA Pulse Scheme for the Backbone Assignment of 15N, 13C, 2H-Labeled Proteins: Application to a 37-kDa Trp Repressor-DNA Complex. *J Am Chem Soc*. 1994;16(22):6464-6465.
36. Nietlispach D, Clowes RT, Broadhurst RW, et al. An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjunction with Random Fractional Deuteration. *J Am Chem Soc*. 1996;118:407-415.
37. Farmer BTI, Venters RA. Assignment of Side-Chain 13C Resonances in Perdeuterated Proteins. *J Am Chem Soc*. 1995;117:4187-4188.
38. Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE. Selective methyl group protonation of perdeuterated proteins. *J Mol Biol*. 1996;263(5):627-636. doi:10.1006/jmbi.1996.0603.
39. Gardner KH, Rosen MK, Kay LE. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry*. 1997;36(6):1389-1401. doi:10.1021/bi9624806.
40. Gardner KH, Kay LE. Production and Incorporation of 15N, 13C, 2H (1H-d1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J Am Chem Soc*. 1997;119:7599-7600.
41. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val , Leu , Ile (δ1). *J Biomol NMR*. 1999;13:369-374.
42. Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP. Exponential Sampling , an Alternative Method for Sampling in Two-Dimensional NMR Experiments. *J Magn Reson*. 1987;73:69-77.
43. Rovnyak D, Sarcone M, Jiang Z. Sensitivity enhancement for maximally

resolved two-dimensional NMR by nonuniform sampling. *Magn Reson Chem*. 2011;49:483-491. doi:10.1002/mrc.2775.

44. Frueh DP. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog Nucl Magn Reson Spectrosc*. 2013;78:47-75. doi:10.1016/j.pnmrs.2013.12.001.
45. Hyberts SG, Takeuchi K, Wagner G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc*. 2010;132(7):2145-2147. doi:10.1021/ja908004w.
46. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR*. 2012;52(4):315-327. doi:10.1007/s10858-012-9611-z.
47. Hoch JC, Maciejewski MW, Filipovic B. Randomization improves sparse sampling in multidimensional NMR. *J Magn Reson*. 2008;193(2):317-320. doi:10.1016/j.jmr.2008.05.011.
48. Maciejewski MW, Mobli M, Schuyler AD, Stern AS, Hoch JC. Data Sampling in Multidimensional NMR: Fundamentals and Strategies. *Top Curr Chem*. 2012;316(1):49-78. doi:10.1007/128.
49. Coggins BE, Venters R a, Zhou P. Radial sampling for fast NMR: Concepts and practices over three decades.Coggins, B. E., Venters, R. a, & Zhou, P. (2010). Radial sampling for fast NMR: Concepts and practices over three decades. Progress in Nuclear Magnetic Resonance Spectroscopy, 57(4), . *Prog Nucl Magn Reson Spectrosc*. 2010;57(4):381-419. doi:10.1016/j.pnmrs.2010.07.001.
50. Hyberts SG, Arthanari H, Robson S a, Wagner G. Perspectives in magnetic resonance: NMR in the post-FFT era. *J Magn Reson*. 2014;241:60-73. doi:10.1016/j.jmr.2013.11.014.
51. Coggins BE, Venters RA, Zhou P. Radial sampling for fast NMR: Concepts and practices over three decades. *Prog Nucl Magn Reson Spectrosc*. 2010;57(4):381-419. doi:10.1016/j.pnmrs.2010.07.001.
52. Farmer BTI. Simultaneous [13C,15N]-HMQC, A Psuedo-Triple-Resonance Experiment. *J Magn Reson*. 19991;93(1):635-641.
53. Boelens R, Burgering M, Fogh RH, Kaptein R. Time-saving methods for heteronuclear multidimensional NMR of (13C, 15N) doubly labeled proteins. *J Biomol NMR*. 1994;4(2):201-213. doi:10.1007/BF00175248.
54. Sattler M, Maurer M, Schleucher J, Griesinger C. A simultaneous 15N,1H- and 13C,1H-HSQC with sensitivity enhancement and a heteronuclear gradient echo. *J Biomol NMR*. 1995;5:97-102. doi:10.1007/BF00227475.
55. Pervushin K, Riek R, Wider G, Wu K. Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in 13 C-Labeled Proteins. 1998;6(9):6394-6400.
56. Lee D, Hilty C, Wider G, Wüthrich K. Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson*. 2006;178:72-76. doi:10.1016/j.jmr.2005.08.014.
57. Frueh D, Arthanari H, Koglin A, Vosburg D. Dynamic thiolation – thioesterase

structure of a non-ribosomal peptide synthetase. *Nature*. 2008;454(August). doi:10.1038/nature07162.

58. Tugarinov V, Hwang PM, Kay LE. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem*. 2004;73:107-146. doi:10.1146/annurev.biochem.73.011303.074004.
59. Rovnyak D, Hoch JC, Stern a S, Wagner G. Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *J Biomol NMR*. 2004;30(1):1-10. doi:10.1023/B:JNMR.0000042946.04002.19.
60. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. 2009;44(4):213-223. doi:10.1007/s10858-009-9333-z.TALOS.
61. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (??1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J Biomol NMR*. 1999;13(4):369-374. doi:10.1023/A:1008393201236.
62. Ayala I, Sounier R, Usé N, Gans P, Boisbouvier J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR*. 2009;43(2):111-119. doi:10.1007/s10858-008-9294-7.
63. Velyvis A, Ruschak AM, Kay LE. An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One*. 2012;7(9):e43725. doi:10.1371/journal.pone.0043725.
64. Fischer M, Kloiber K, Häusler J, Ledolter K, Konrat R, Schmid W. Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *ChemBioChem*. 2007;8(6):610-612. doi:10.1002/cbic.200600551.
65. Weininger U, Liu Z, Mcintyre DD, Vogel HJ, Akke M. Specific 12CbD212CcD2S13CeHD2 isotopomer labeling of methionine to characterize protein Dynamics by 1H and 13C NMR relaxation dispersion. *J Am Chem Soc*. 2012;134:18562-18565.
66. Hilty C, Fernández C, Wider G, Wüthrich K. Side chain NMR assignments in the membrane protein OmpX reconstituted in DHPC micelles. *J Biomol NMR*. 2002;23(4):289-301.
67. Tugarinov V, Kay LE. Side chain assignments of Ile delta 1 methyl groups in high molecular weight proteins: an application to a 46 ns tumbling molecule. *J Am Chem Soc*. 2003;125(19):5701-5706. doi:10.1021/ja021452+.
68. Tugarinov V, Kay LE. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc*. 2003;125(45):13868-13878. doi:10.1021/ja030345s.
69. Sheppard D, Guo C, Tugarinov V. 4D 1H-13C NMR spectroscopy for assignments of alanine methyls in large and complex protein structures. *J Am Chem Soc*. 2009;131(4):1364-1365. doi:10.1021/ja808202q.
70. Sheppard D, Guo C, Tugarinov V. Methyl-detected “out-and-back” NMR experiments for simultaneous assignments of Alabeta and Ilegamma2 methyl groups in large proteins. *J Biomol NMR*. 2009;43(4):229-238. doi:10.1007/s10858-009-9305-3.

71. Guo C, Tugarinov V. Selective 1H- 13C NMR spectroscopy of methyl groups in residually protonated samples of large proteins. *J Biomol NMR*. 2010;46(2):127-133. doi:10.1007/s10858-009-9393-0.
72. Frueh DP, Vosburg DA, Walsh CT, Wagner G. Determination of all nOes in 1H-13C-Me-ILV-U-2H-15N proteins with two time-shared experiments. *J Biomol NMR*. 2006;34(1):31-40. doi:10.1007/s10858-005-5338-4.
73. Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc*. 2005;127(8):2767-2775. doi:10.1021/ja044032o.
74. Coggins BE, Werner-allen JW, Yan A, Zhou P. Rapid Protein Global Fold Determination Using Ultrasparse Sampling, High-Dynamic Range Artifact Suppression, and Time- Shared NOESY. *J Am Chem Soc*. 2012;134:18619-18630.
75. Mishra SH, Harden BJ, Frueh DP. A 3D time-shared NOESY experiment designed to provide optimal resolution for accurate assignment of NMR distance restraints in large proteins. *J Biomol NMR*. 2014;60(4):265-274. doi:10.1007/s10858-014-9873-8.
76. Bermel W, Felli IC, Kümmerle R, Pierattelli R. 13C Direct-Detection Biomolecular NMR. *Concepts Magn Reson*. 2008;32A(3):183-200. doi:10.1002/cmr.a.
77. Shimba N, Stern AS, Craik CS, Hoch JC, Dötsch V. Elimination of 13Calpha splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. *J Am Chem Soc*. 2003;125(9):2382-2383. doi:10.1021/ja027973e.
78. Jordan JB, Kovacs H, Wang Y, et al. Protonated Proteins : Facile Methyl Resonance Assignment and Protein Structure Determination. 2006;(4):12366-12371.
79. Duma L, Hediger S, Lesage A, Emsley L. Spin-state selection in solid-state NMR. *J Magn Reson*. 2003;164(1):187-195. doi:10.1016/S1090-7807(03)00187-3.
80. Andersson P, Weigelt J, Otting G. Spin-state selection filters for the measurement of heteronuclear one-bond coupling constants. *J Biomol NMR*. 1998;12(3):435-441. http://www.ncbi.nlm.nih.gov/pubmed/9835050.
81. Ottiger M, Delaglio F, Bax a. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson*. 1998;131(2):373-378. doi:10.1006/jmre.1998.1361.
82. Shimba N, Kovacs H, Stern AS, et al. Optimization of 13C direct detection NMR methods. *J Biomol NMR*. 2004;30(2):175-179. doi:10.1023/B:JNMR.0000048855.35771.11.
83. Nielsen NC, Thøgersen H, Sørensen OW. Doubling the Sensitivity of INADEQUATE for Tracing Out the Carbon Skeleton of Molecules by. *J Am Chem Soc*. 1995;117:11365-11366.
84. Bermel W, Bertini I, Felli I, Piccioli M, Pierattelli R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Prog Nucl Magn Reson Spectrosc*. 2006;48(1):25-45. doi:10.1016/j.pnmrs.2005.09.002.

85. Serber Z, Richter C, Dötsch V. Carbon-detected NMR experiments to investigate structure and dynamics of biological macromolecules. *Chembiochem*. 2001;2(4):247-251. http://www.ncbi.nlm.nih.gov/pubmed/11828451.
86. Serber Z, Richter C, Moskau D, et al. New Carbon-Detected Protein NMR Experiments Using CryoProbes. *J Am Chem Soc*. 2000;122:3554-3555.
87. Bermel W, Bertini I, Felli I. Speeding up 13C direct detection Biomolecualr NMR spectroscopy. *J Am Chem Soc*. 2009;(31):15339-15345.
88. Pervushin K, Eletsky A. A new strategy for backbone resonance assignment in large proteins using a MQ-HACACO experiment. *J Biomol NMR*. 2003;25(2):147-152. http://www.ncbi.nlm.nih.gov/pubmed/12652123.
89. Fischer MWF, Zeng L, Zuiderweg ERP, June R V. Use of 13 C - 13 C NOE for the Assignment of NMR Lines of Larger Labeled Proteins at Larger Magnetic Fields. 1996;7863(9):12457-12458.
90. Eletsky A, Moreira O, Kovacs H, Pervushin K. A novel strategy for the assignment of side-chain resonances in completely deuterated large proteins using 13C spectroscopy. *J Biomol NMR*. 2003;26(2):167-179. http://www.ncbi.nlm.nih.gov/pubmed/12766412.
91. Doddrell D., Pegg D., Bendall M. Distortionless enhancement of NMR signals by polarization transfer. *J Magn Reson*. 1982;48(2):323-327. doi:10.1016/0022-2364(82)90286-4.
92. Loria JP, Rance M, Palmer a G. Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J Magn Reson*. 1999;141(1):180-184. doi:10.1006/jmre.1999.1891.
93. Eletsky a, Kienhöfer a, Pervushin K. TROSY NMR with partially deuterated proteins. *J Biomol NMR*. 2001;20(2):177-180. http://www.ncbi.nlm.nih.gov/pubmed/11495249.
94. Bertini I, Felli IC, Kümmerle R, Moskau D, Pierattelli R. 13C-13C NOESY: an attractive alternative for studying large macromolecules. *J Am Chem Soc*. 2004;126(2):464-465. doi:10.1021/ja0357036.
95. Bertini I, Felli IC, Kümmerle R, Luchinat C, Pierattelli R. 13C-13C NOESY: a constructive use of 13C-13C spin-diffusion. *J Biomol NMR*. 2004;30(3):245-251. doi:10.1007/s10858-005-1679-2.
96. Hu K, Vögeli B, Pervushin K. Side-chain H and C resonance assignment in protonated/partially deuterated proteins using an improved 3D(13)C-detected HCC-TOCSY. *J Magn Reson*. 2005;174(2):200-208. doi:10.1016/j.jmr.2005.02.008.
97. Shekhtman A, Ghose R, Goger M, Cowburn D. NMR structure determination and investigation using a reduced proton (REDPRO) labeling strategy for proteins. *FEBS Lett*. 2002;524:177-182.
98. Liao X, Tugarinov V. Selective detection of $^{13}CHD_2$ signals from a mixture of $^{13}CH_3$/$^{13}CH_2D$/$^{13}CHD_2$ methyl isotopomers in proteins. *J Magn Reson*. 2011;209(1):101-107. doi:10.1016/j.jmr.2010.12.014.
99. Finking R, Marahiel M a. Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol*. 2004;58:453-488. doi:10.1146/annurev.micro.58.030603.123615.
100. Mootz HDf, Schwarzer D, Marahiel M a. Ways of assembling complex natural

products on modular nonribosomal peptide synthetases. *Chembiochem*. 2002;3(6):490-504. doi:10.1002/1439-7633(20020603)3:6<490::AID-CBIC490>3.0.CO;2-N.

101. Marahiel M a., Stachelhaus T, Mootz HD. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev*. 1997;97(7):2651-2674. http://www.ncbi.nlm.nih.gov/pubmed/11851476.
102. Fischbach M a, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev*. 2006;106(8):3468-3496. doi:10.1021/cr0503097.
103. Keating T a, Miller D a, Walsh CT. Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase. *Biochemistry*. 2000;39(16):4729-4739. http://www.ncbi.nlm.nih.gov/pubmed/10769129.
104. Gehring A, Mori I, Perry R, Walsh C. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of yersinia pestis. *Biochemistry*. 1998;37(48):17104. doi:10.1021/bi9850524.
105. Gehring a M, DeMoll E, Fetherston JD, et al. Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by Yersinia pestis. *Chem Biol*. 1998;5(10):573-586. http://www.ncbi.nlm.nih.gov/pubmed/9818149.
106. Patel HM, Tao J, Walsh CT. Epimerization of an L-cysteinyl to a D-cysteinyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from Pseudomonas aeruginosa. *Biochemistry*. 2003;42(35):10514-10527. doi:10.1021/bi034840c.
107. Perry RD, Fetherston JD. Yersiniabactin iron uptake: mechanisms and role in Yersinia pestis pathogenesis. *Microbes Infect*. 2011;13(10):808-817. doi:10.1016/j.micinf.2011.04.008.
108. Parkhill J, Wren BW, Thomson NR, et al. Genome sequence of Yersinia pestis , the causative agent of plague. *Nature*. 2003;413(2001):523-527.
109. Keating T a, Suo Z, Ehmann DE, Walsh CT. Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. *Biochemistry*. 2000;39(9):2297-2306. http://www.ncbi.nlm.nih.gov/pubmed/10694396.
110. Youard ZA, Reimmann C. Stereospecific recognition of pyochelin and enantio-pyochelin by the PchR proteins in fluorescent pseudomonads. *Microbiology*. 2010;156(Pt 6):1772-1782. doi:10.1099/mic.0.037796-0.
111. Brillet K, Reimmann C, Mislin GL a, et al. Pyochelin enantiomers and their outer-membrane siderophore transporters in fluorescent pseudomonads: structural bases for unique enantiospecific recognition. *J Am Chem Soc*. 2011;133(41):16503-16509. doi:10.1021/ja205504z.
112. Hoegy F, Lee X, Noel S, et al. Stereospecificity of the siderophore pyochelin outer membrane transporters in fluorescent pseudomonads. *J Biol Chem*. 2009;284(22):14949-14957. doi:10.1074/jbc.M900606200.
113. Bracher PJ, Snyder PW, Bohall BR, Whitesides GM. The relative rates of thiol-thioester exchange and hydrolysis for alkyl and aryl thioalkanoates in

water. *Orig Life Evol Biosph*. 2011;41(5):399-412. doi:10.1007/s11084-011-9243-4.

114. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000;16(4):404-405. doi:10.1093/bioinformatics/16.4.404.
115. Kelley LA, Mezulis S, Yates C, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat Protoc*. 2015;10(6):845-858. doi:10.1038/nprot.2015-053.
116. Gelly J-C, Etchebest C, Hazout S, de Brevern a G. Protein Peeling 2: a web server to convert protein structures into series of protein units. *Nucleic Acids Res*. 2006;34(Web Server issue):W75-W78. doi:10.1093/nar/gkl292.
117. Cornilescu G, Delaglio F, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR*. 1999;13:289-302. doi:10.1007/s10858-007-9166-6.
118. Shen Y, Bax A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR*. 2013;56(3):227-241. doi:10.1007/s10858-013-9741-y.
119. Sundlov JA, Shi C, Wilson DJ, Aldrich CC, Gulick AM. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem Biol*. 2012;19(2):188-198. doi:10.1016/j.chembiol.2011.11.013.
120. Kochan G, Pilka ES, von Delft F, Oppermann U, Yue WW. Structural Snapshots for the Conformation-dependent Catalysis by Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A. *J Mol Biol*. 2009;388(5):997-1008. doi:10.1016/j.jmb.2009.03.064.
121. Conti E, Stachelhaus T, Marahiel M a, Brick P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J*. 1997;16(14):4174-4183. doi:10.1093/emboj/16.14.4174.
122. Reger AS, Wu R, Dunaway-Mariano D, Gulick AM. Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry*. 2008;47(31):8016-8025. doi:10.1021/bi800696y.
123. Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA, Stubbs MT. Crystal structure of DltA: Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J Biol Chem*. 2008;283(47):32484-32491. doi:10.1074/jbc.M800557200.
124. Hu Y, Gai Y, Yin L, et al. Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. *Plant Cell*. 2010;22(9):3093-3104. doi:10.1105/tpc.109.072652.
125. Lee TV, Johnson LJ, Johnson RD, et al. Structure of a eukaryotic nonribosomal peptide synthetase adenylation domain that activates a large hydroxamate amino acid in siderophore biosynthesis. *J Biol Chem*. 2010;285(4):2415-2427. doi:10.1074/jbc.M109.071324.
126. Grzesiek S, Bax a. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR*. 1993;3(2):185-204. http://www.ncbi.nlm.nih.gov/pubmed/8477186.

127. Brüschweiler R, Zhang F. Covariance nuclear magnetic resonance spectroscopy. *J Chem Phys*. 2004;120(11):5253-5260. doi:10.1063/1.1647054.
128. Trbovic N, Smirnov S, Zhang F, Br??schweiler R. Covariance NMR spectroscopy by singular value decomposition. *J Magn Reson*. 2004;171(2):277-283. doi:10.1016/j.jmr.2004.08.007.
129. Zhang F, Brüschweiler R. Indirect covariance NMR spectroscopy. *J Am Chem Soc*. 2004;126(41):13180-13181. doi:10.1021/ja047241h.
130. Short T, Alzapiedi L, Brüschweiler R, Snyder D. A covariance NMR toolbox for MATLAB and OCTAVE. *J Magn Reson*. 2011;209(1):75-78. doi:10.1016/j.jmr.2010.11.018.
131. Kupče E, Freeman R. Hyperdimensional NMR Spectroscopy. *J Am Chem*. 2006;(i):6020-6021. http://pubs.acs.org/doi/abs/10.1021/ja0609598. Accessed July 6, 2012.
132. Benison G, Berkholz DS, Barbar E. Protein assignments without peak lists using higher-order spectra. *J Magn Reson*. 2007;189(2):173-181. doi:10.1016/j.jmr.2007.09.009.
133. Chen K, Delaglio F, Tjandra N. A practical implementation of cross-spectrum in protein backbone resonance assignment. *J Magn Reson*. 2010;203(2):208-212. doi:10.1016/j.jmr.2009.12.018.
134. Lescop E, Brutscher B. Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J Am Chem Soc*. 2007;129(39):11916-11917. doi:10.1021/ja0751577.
135. Brüschweiler R. Theory of covariance nuclear magnetic resonance spectroscopy. *J Chem Phys*. 2004;121(1):409-414. doi:10.1063/1.1755652.
136. Short T, Alzapiedi L, Brüschweiler R, Snyder D. A covariance NMR toolbox for MATLAB and OCTAVE. *J Magn Reson*. 2011;209(1):75-78. doi:10.1016/j.jmr.2010.11.018.
137. *Matlab Reference Guide*. Natick, MA; 1992.
138. Harden BJ, Mishra SH, Frueh DP. Effortless assignment with 4D covariance sequential correlation maps. *J Magn Reson*. 2015;260:83-88. doi:10.1016/j.jmr.2015.09.007.
139. Gardner KH, Kay LE. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct*. 1998;27:357-406. doi:10.1146/annurev.biophys.27.1.357.
140. Goto NK, Kay LE. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol*. 2000;10(5):585-592. http://www.ncbi.nlm.nih.gov/pubmed/11042458.
141. Tugarinov V, Kay LE. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J Am Chem Soc*. 2003;125(45):13868-13878. doi:10.1021/ja030345s.
142. Shen Y, Bax A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR*. 2013;56:227-241. doi:10.1007/s10858-013-9741-y.
143. Isaacson RL, Simpson PJ, Liu M, et al. A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc*. 2007;129(50):15428-15429. doi:10.1021/ja0761784.

144. Krejcirikova A, Tugarinov V. 3D-TROSY-based backbone and ILV-methyl resonance assignments of a 319-residue homodimer from a single protein sample. *J Biomol NMR*. 2012;54(2):135-143. doi:10.1007/s10858-012-9667-9.
145. Marino JP, Diener JL, Moore PB, Griesinger C. Multiple-Quantum Coherence Dramatically Enhances the Sensitivity of CH and CH 2 Correlations in Uniformly 13 C-Labeled RNA. 1997;7863(3):7361-7366.
146. Ollerenshaw JE, Tugarinov V, Kay LE. Methyl TROSY: explanation and experimental verification. *Magn Reson Chem*. 2003;41(10):843-852. doi:10.1002/mrc.1256.
147. Mishra SH, Frueh DP. Assignment of methyl NMR resonances of a 52 kDa protein with residue-specific 4D correlation maps. *J Biomol NMR*. 2015;62(3):281-290. doi:10.1007/s10858-015-9943-6.
148. Cai M, Huang Y, Sakaguchi K, Clore GM, Gronenborn a M, Craigie R. An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli. *J Biomol NMR*. 1998;11(1):97-102. doi:10.1023/A:1008222131470.
149. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. *J Biomol NMR*. 1995;6(3):277-293.
150. Roehrl MH a, Heffron GJ, Wagner G. Correspondence between spin-dynamic phases and pulse program phases of NMR spectrometers. *J Magn Reson*. 2005;174(2):325-330. doi:10.1016/j.jmr.2005.02.001.
151. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *JBiomolNMR*. 1992;2:661-666.
152. Shaka AJ, Reeler J, Frenkiel T, Freeman R. An Improved Sequence for Broadband Decoupling : WALTZ-16. *J Magn Reson*. 1983;52:335-338.
153. van Ingen H, Vuister GW, Tessari M. A Two-Dimensional Artifact from Asynchronous Decoupling. *J Magn Reson*. 2002;156(2):258-261. doi:10.1006/jmre.2002.2564.
154. Marion D, Wüthrich K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for mreasurement of 1H-1H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun*. 1983;113(3):967-974. doi:10.1017/CBO9781107415324.004.

Curriculum Vitae for Ph.D. Candidates

The Johns Hopkins University School of Medicine

Scott Robert Nichols
Place of Birth: Omaha, Nebraska

Education

Johns Hopkins School of Medicine, Baltimore, MD June 2016 (expected)
Ph.D. Molecular Biophysics
Adviser: Dr. Dominique Frueh
Thesis title: "NMR characterization of the Yersiniabactin Synthetase Epimerization Domain"

University of Iowa, Iowa City, IA 2003-2007
B.S. Honors in Biochemistry
Adviser: Dr. Theresa Gioannini
Thesis Title: Using *galE Escherichia coli* to study effects of LPS sugar chain length on interactions of LPS with innate immune proteins

Research Experience

Johns Hopkins University School of Medicine, Baltimore, MD
June 2010 – present
Graduate research with Dr. Dominique Frueh

Brown University, Providence, RI
August 2008—May 2010
Research assistant with Dr. Wolfgang Peti

Pasteur Institute, Paris, France
September 2007 – May 2008
Fulbright Advanced Student with Dr. Lhousseine Touqui

University of Iowa, Iowa City, IA
May 2006—September 2007
Honors undergraduate research with Dr. Theresa Gioannini

University of Iowa, Iowa City, IA
January 2007—May 2007
Undergraduate research with Dr. Adrian Elcock

Awards & Recognition

American Heart Association Pre-doctoral Fellowship Recipient	2014-2016
Finalist Johns Hopkins Graduate Consulting Club Mini-Case Competition	2015
National Science Foundation Graduate Research Fellowship Program Honorable Mention	2012
Hopkins Francis Carlson Fellow	2010-2012
Fulbright Laureate	2007-2008
University of Iowa Research Experience for Undergradautes Awardee	2006-2007
University of Iowa Kathleen K. Beninga Scholarship in Microbiology	2005-2006

Presentations

NMR characterization of an unusual 37 kDa epimerization enzyme
Biophysical Society Annual Meetings: 2011 (Baltimore, MD), 2013 (Philadelphia, PA), 2015 (Baltimore, MD)

Modeling a complete NRPS module of Yersiniabactin Synthetase
Workshop on Hydrogdynamics and Thermodynamic Analysis of Macromolecules: 2011 (Bethesda, MD NIH)

Regulation of Protein Phosphatase 1
Keystone Symposium on Structural Biology: 2010 (Breckenridge, CO) &
Protein Society Annual Meeting: 2009 (Boston, MA)

Generation and characterization of an acetate auxotroph of *galE E. coli*: a tool to LPS interactions with endotoxin-binding proteins
International Endotoxin and Innate Immunity Society Joint Meeting: 2006 (San Antonio, TX)

"Rough v. Smooth LPS and the Innate Immune System"
Iowa State University Iowa State University Stupka Microsymposium: 2007 (Ames, IA)

Teaching Experience

Department of Biophysics & Biophysical Chemistry, Johns Hopkins University School of Medicine

Exam Grader, Biochemical & Biophysical Principles	2012-2015
Mentor to Master's Exchange Student	2013
Teaching Assistant, Using Structure to Understand Biology	2015

Department of Biophysics, Johns Hopkins University

Teaching Assistant, Methods in Molecular Biophysics 2012

Leadership Experience

Department of Biophysics, Johns Hopkins University

Biomolecular NMR Journal Club Facilitator 2013

Department of Biophysics & Biophysical Chemistry, Johns Hopkins University School of Medicine

Student-invited Speaker Coordinator 2012-2014

Publications

Harden BJ*, **Nichols SR***, Frueh DP "Facilitated assignment of large protein NMR signals with covariance sequential spectra using spectral derivatives," *Journal of the American Chemical Society*, 2014

Goodrich AC*, **Nichols SR***, Frueh DP "Resonance Assignments and Structure Determination of Large and Challenging Proteins," *Encyclopedia of NMR*, 2014

Frueh DP, Goodrich AC, Mishra SH, **Nichols SR** "NMR Methods for Structural Studies of Large Monomeric and Multimeric Proteins," *Current Opinion in Structural Biology*, 2013

O'Connell N*, **Nichols SR***, Heroes E, Beullens M, Bollen, M, Peti W, Page R "The Molecular Basis for Substrate Specificity of the Nuclear NIPP:PP1 Holoenzyme," *Structure*, 2012

*denotes equal contribution to authorship